SOUTH AFRICAN
Journal of Science

SEPTEMBER/OCTOBER 2016

eISSN: 1996-7489

Applying scientific thinking in the service of society

20 YEARS

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

volume 112
number 9/10

SEPTEMBER/OCTOBER

eISSN: 1996-7489

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
Sue Randall

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

SOUTH AFRICAN

Journal of Science

volume 112

number 9/10

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) with financial assistance from the Department of Science & Technology.

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

Correspondence and enquiries
sajs@assaf.org.za

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

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

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

Cover caption
Pteronia camphorata – an aromatic shrub endemic to the western and southern coastal region of South Africa – identified for the first time by Hulley and colleagues (photo: B-E van Wyk). *Pteronia camphorata* is an important medicinal plant known as |*nhora* to the San and Nama people. In an article on page 84, Hulley and colleagues validate the traditional uses of this plant.

Review Article

Research Article

No fee increases: Cascades and calamities

So much has been written, over the past 11 months, about '#feesmustfall' and '#nofees', that there is little, if anything, left to say. The Council on Higher Education has presented its report to the Minister of Higher Education; the Presidential Commission continues with its work; and Universities South Africa has made representations to the Minister and to the Parliamentary Portfolio Committee on Higher Education. All that is needed is a decision (now arrogated from university councils it would seem) – and the means of dealing with whatever will follow the decision.

However, in the welter of the situation faced by the South African education system, and on the implications of the decisions that need to be made, two issues seem to have been overlooked.

The first of these is what might be called 'the cascade effect'; the second is the calamity of the discarded.

Despite denials and various alternative explanations, it remains true that, on 23 October 2015, President Zuma announced that there would be no fee increases for university students in 2016. The consequences and dangers of this announcement are well known – but it was not, then, immediately obvious (although perhaps it ought to have been) that the demand for no fee increases so readily acceded to would very quickly transmogrify into 'free higher education for all'. Economic impossibility and social justice irrationality notwithstanding, both issues now lie at the core of the problem.

Yet there is a significant implication that is not often addressed. Just over 1 million students are enrolled in the higher education system in South Africa. That they *are* enrolled makes them a small but privileged part of society: they are about 1% of the young people who entered primary school 12 years ago. Setting that privilege aside (while not ignoring the reality that as few as 50–60% of those students will eventually graduate), the demand that there should be not only no fee increases but, possibly, no fees at all, leads to a rather different view of the situation.

In this context, and that of current and unexpected political uncertainty, consider what might happen if a 'no fee increase' or a 'no fees at all' demand were to be met. Irrespective of how socially unjust it is (the wealthy will be the greatest beneficiaries of a 'no fee' higher education system), it will have serious consequences for the quality of South African university education. Egregious though that may be, it will not be the worst scenario for the South African education system. The international rankings of our universities have already shown a lowering of the status of South African universities – a result of, according to some experts, the persistent underfunding of higher education by the South African government. This is, however, just a start of the challenges that the South African post-school education system must be prepared to face.

If the most privileged young people should become entitled to low-cost or free higher education, why would the (currently) rather less privileged 800 000 or so students in the Technical and Vocational Education Training (TVET) sector not be entitled – even more entitled – to free education?

The cascade begins.

If, then, the more and most privileged pay fixed, or no, fees, why then should the most deserving – the young people at school – pay anything at all? The cascade ends with low cost or free education for all, while the economy faces a growth rate close to zero – a situation that will continue until more well-educated and skilled young people enter and contribute to that same economy.

Some observers and interpreters have proposed that the student protests, burnings and broader mayhem, have to do with higher education's callous alienation of the relatively few young people who are enrolled as students. These observers do not, of course, subscribe to the idea that young people who enter the higher education system are privileged. Nor, for the most part, do they acknowledge the predominant class of the protesters, who speciously claim to be the representatives and champions of the poor. Claiming 'no increases', or free higher education, is not a plea for the poor – it is a hollow middle-class avoidance of one of the most critical issues: the need for comprehensive support for smart, deserving young people who need, not just their fees, but also their living to be assured.

Which leads to the calamity. Which will almost certainly turn out to be more serious.

All told, there are some 2 million learners in the university, TVET and community college sector. Yet, in the same age group (18–25), there are over 3 million young people who are NEETs – 'not in employment, education or training'.

These are the 3 million young people who are adrift and seemingly without options – and who do not receive, amidst the fees debate, the attention or the financial support that they deserve. They beg for food at street corners; they become, out of desperation, members of gangs; or else, in the most dire circumstances, they are already in prison. Of their conditions, the self-proclaimed socio-economic warriors are either unforgivably ignorant or, equally unforgivably, careless.

If ever there was a clear insight into the real values of protesting students (many, the sons and daughters of chief executive officers, vice chancellors and Members of Parliament), then it is this – that they ignore, or at the very least overlook, the severe and long-term social and economic implications of the NEETs. Free education, or education without fee increases, is not about social justice. It can only be a very sad statement about ignorance, carelessness of social and economic realities, and the promotion of some other, yet to be revealed (and possibly self-serving), end.

What does all this say about universities, in these difficult times? Perhaps two things: they clearly do not do enough to teach vital, core values; and some of their leaders have abandoned academic principles in favour of expedience when principle is most needed.

HOW TO CITE: Butler-Adam J. No fee increases: Cascades and calamities. S Afr J Sci. 2016;112(9/10), Art. #a0176, 1 page. http://dx.doi.org/10.17159/sajs.2016/a0176

The Golden Ratio (1.62) as a dimensionless biological constant

AUTHOR:
Michael Lieber[1]

AFFILIATION:
[1]Previous: United States Environmental Protection Agency, Division 9, San Francisco, California, USA

CORRESPONDENCE TO:
Michael Lieber

EMAIL:
michaellieber@juno.com

KEYWORDS:
unifying role; general relativity; Einstein; space-time

HOW TO CITE:
Lieber M. The Golden Ratio (1.62) as a dimensionless biological constant. S Afr J Sci. 2016;112(9/10), Art. #a0172, 1 page. http://dx.doi.org/10.17159/sajs.2016/a0172

Regarding the significance of the Golden Ratio, 1.62, in phenomena as described by Boeyens and Thackeray[1] in 2014, I also – in a 1998 publication[2] – demonstrated its significant and unifying role in physical and biological reality. It was demonstrated that the Golden Ratio is an inherent, dimensionless biological constant composing the dimensional physical constants of physics. In that article, it was also shown that this biological constant reflects a regenerative, adaptive or accommodating process operating through all scales of reality, from the sub-quantum, through the biological, to the cosmological level of organisation, and in so doing, giving a fractal unity to such realities, uniting quantum reality with cosmological reality, hence quantum reality with the reality of general relativity. Through a simple mathematical analysis, it was also found and noted that this dimensionless biological constant denotes a near infinitesimal, regenerative feature of space-time. Also described was how the dimensionless biological constant defined reality or space-time as having a vortical or spiral morphology. In 1949, the mathematician Kurt Godel, in solving in a new way the differential field equations from general relativity, showed the universe as having a spiral or vortical space-time geometry through the rotation of matter.[3]

In a subsequent article, published online in February 2015 (www.michaellieber.com), I further elaborated on the role of the Golden Ratio as a biological constant in structuring regeneratively and self-similarly the different levels of reality. Among the many related subjects covered, it was shown how this constant might have united the general relativity theory of Einstein with the innovative quantum mechanics described by the physicist Paul Dirac. From the standpoint of biology, it was also noted how this biological constant reflects processes in evolution. In viewing these two articles, it becomes clear how this dimensionless biological constant reflects and defines an underlying unity within reality.

The findings of Boeyens and Thackeray[1] give significant support to the conclusions and implied predictions that I presented in 1998. It is encouraging that others are pursuing and describing the significant role that the Golden Ratio has in unifying universal processes and states, thereby giving us a deeper, holistic understanding of our universe. Such may lead to a testable, universal view of reality – perhaps to a type of biological, unified, field theory.

References

1. Boeyens JCA, Thackeray JF. Number theory and the unity of science. S Afr J Sci. 2014;110(11/12), Art. #a0084, 2 pages. http://dx.doi.org/10.1590/sajs.2014/a0084
2. Lieber M. The living spiral. A dimensionless biological constant gives a new perspective to physics. Rivista di Biologia – Biology Forum. 1998;91:91–118.
3. Gödel K. An example of a new type of cosmological solutions of Einstein's field equations of gravitation. Rev Mod Phys. 1949;21(3):447. http://dx.doi.org/10.1103/RevModPhys.21.447

Source: Boeyens and Thackeray[1]

Examples of the Golden Ratio found in nature: (from left to right) the Whirlpool Galaxy, a Nautilus shell, Hurricane Katrina and an ammonite.

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

National Research Foundation celebrates science excellence for development

AUTHOR:
Thalia Brussow[1]

AFFILIATION:
[1]National Research Foundation, Pretoria, South Africa

CORRESPONDENCE TO:
Thalia Brussow

EMAIL:
thalia.brussow@nrf.ac.za

KEYWORDS:
National System of Innovation; knowledge creation; research excellence; South Africa

HOW TO CITE:
Brussow T. National Research Foundation celebrates science excellence for development. S Afr J Sci. 2016;112(9/10), Art. #a0178, 3 pages. http://dx.doi.org/10.17159/sajs.2016/a0178

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

Knowledge is the currency of a thriving economy. Generating and engaging with that knowledge delivers innovative ideas that empower and sustain nations. That is why the National Research Foundation (NRF) of South Africa annually celebrates the work of our country's finest researchers through the NRF awards.

The NRF awards recognise South African scientists for their achievements as indicated through the NRF rating system. In addition to the rating-linked awards, special recognition awards provide a platform to honour researchers for career achievements and contributions to knowledge creation and dissemination, as well as for capacity development and transformation in the National System of Innovation. This year's awards ceremony took place on 1 September 2016 in Limpopo.

The men and women celebrated through the awards are in various stages of their research development – some have shown outstanding academic performance in their final year of PhD study, while others have been contributing to the National System of Innovation for the bigger part of their careers and are deemed to be international leaders in their fields. Yet despite their differences, this cohort has something very important in common – without their valuable work, we would not be able to advance humanity's body of knowledge or secure a brighter future for all South Africans.

Speaking at the awards the Minister of Science and Technology, Mrs Naledi Pandor, noted that South Africa has a scientific heritage to be proud of and that 'We have made a disproportionate contribution to scientific research worldwide, considering our relatively small economy'.

Minister Pandor further observed that South Africa faces a challenge when it comes to training and retaining scientists and researchers:

> *Significant investment is required in both human development and research infrastructure at universities both to increase the productivity of the system and to support the development of research capacity at formerly black universities and universities of technology.*
>
> *Our challenge is to address gender and racial imbalances in the make-up of our science and technology workforce. We not only want to encourage more students to embark on science and engineering studies, but we are also making plans to sustain their ability to pursue research careers. More than ever, international fellowship and bursaries will be important if we are to achieve our target of producing 5000 doctorates per annum, 3000 of which should be in science, engineering and technology fields.*

By recognising the research talent that South Africa has, the NRF hopes to encourage the next generation of scientists to develop their own talents as well as to promote the public's understanding of and engagement with science.

Special Awards

Seven Special Awards were presented to eight individuals and one higher education institution.

Nominations for the Special Awards were received from South African higher education institutions. The nominations were then evaluated by an expert panel which made recommendations to the NRF's Corporate Executive which made the final decision.

For the Excelleration Award, the NRF makes use of data that is supplied by the Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP) to select the winning institution.

Research Excellence Award for Next Generation Researchers

The Research Excellence Award for Next Generation Researchers is awarded to full-time final-year NRF-funded doctoral students (one woman and one man) who have achieved outstanding academic performance, and demonstrated the potential for contributing significantly to the National System of Innovation. This year's recipients were: Dr Pragashnie Govender (Health Sciences, University of KwaZulu-Natal) and Mr Sooraj Baijnath (Pharmaceutical Chemistry, University of KwaZulu-Natal).

Research Excellence Award for Early Career/Emerging Researchers

The Thuthuka funding instrument is central to the NRF's human capital development strategy and aspires to improve the research capacities of designated researchers with the ultimate aim of redressing historical imbalances. The Research Excellence Award for Early Career/Emerging Researchers award recognises two post-PhD Thuthuka grantholders (one woman and one man) who have achieved exceptional research performance. The awardees this year were: Prof. Nosipho Moloto (Chemistry, University of the Witwatersrand) and Prof. Mark Engel (Medical Sciences, University of Cape Town).

Excellence in Science Engagement Award

This award recognises an individual in the research community at a South African higher education institution or science council who has made an outstanding contribution to public engagement with and understanding of various

areas of science over a sustained period. This year the Excellence in Science Engagement Award went to Prof. Lee Berger (Human Evolution, University of the Witwatersrand).

NRF Excelleration Award for South African Research Institutions

Derived from the words 'excellence' and 'acceleration', the Excelleration Award acknowledges South African research institutions for achieving the most improved research performance over recent years, measured against a selection of critical indicators. This year the award went to the University of South Africa. The University of South Africa also received a prize sponsored by Thomson Reuters.

Hamilton Naki Award

This award recognises an individual for outstanding efforts to advance their career in science, and for achieving world-class research performance, despite considerable equity challenges. This award was named after Mr Hamilton Naki, a self-taught surgeon who developed his career against all odds. The recipient of the 2016 award was Prof. Lerothodi Leeuw (Astrophysics, University of South Africa).

Champion of Research Capacity Development and Transformation at South African Higher Education Institutions Award

This award is awarded to individuals within the research community who contribute to the transformation of South Africa's community and landscape.

Dr Pragashnie Govender and Mr Sooraj Baijnath receive the Research Excellence Award for Next Generation Researchers

Prof. Nosipho Moloto and Prof. Mark Engel receive the Research Excellence Award for Early Career/Emerging Researchers

Prof. Lee Berger (middle) is the recipient of the Excellence in Science Engagement Award

Prof. Lerothodi Leeuw (middle) receives the Hamilton Naki Award

Prof. José Frantz (middle) is the recipient of the Champion of Research Capacity Development and Transformation at South African Higher Education Institutions Award

Prof. Chabani Manganyi receives the prestigious Lifetime Achievement Award

The award is dependent on the number of students in the designated groups who had been trained, as well as the quality and impact of research outputs of the students. Prof. José Frantz (Physiotherapy, University of the Western Cape) received this award in 2016.

Lifetime Achievement Award

The NRF's most prestigious award honours a deserving South African individual who is considered to have made outstanding or extraordinary contributions, of international standing and impact, to the development of science in and for South Africa over an extended period of time.

For his contribution as a psychologist, a scholar, a leader in higher education, a key role player in the transformation of the South African education system, a biographer, a gentle intellectual with an enduring love for this country and its people, Prof. Chabani Manganyi was awarded the Lifetime Achievement Award.

In recent years, Prof. Chabani Manganyi has also been acknowledged through honorary doctorates from the University of the Witwatersrand and the University of South Africa, and has received a certificate of acknowledgement from Rhodes University.

NRF ratings-based awards

At the ceremony, awards were also given to researchers who earned A- and P-ratings.

The ratings are awarded through the peer-review based NRF rating system. In order to receive a rating, researchers have to apply and submit evidence of their work. Assessing the work of applicants is a rigorous process that involves a network of peer reviewers, 22 specialist committees, five externally appointed chairpersons and five assessors, the Executive Evaluation Committee and the Appeals Committee.

Ratings are awarded based on the quality of a researcher's work. Reviewers do not count the number of publications that researchers have produced, but rather weigh the outputs through an array of established metrics.

A-rated researchers are unequivocally recognised by their peers as leading international scholars in their respective fields for the high quality and impact of recent research outputs. This year 27 researchers received A-ratings (Table 1).

P-ratings are assigned to researchers under the age of 35, who have held a doctorate or equivalent qualification for less than 5 years at the time of application. These researchers are considered likely to become future international leaders in their respective fields, on the basis of exceptional potential demonstrated in research performance and output. This year P-ratings were awarded to four researchers:

- Dr Katye Altieri, University of Cape Town (Earth and Marine Sciences)
- Dr Robyn Pickering, University of Cape Town (Earth and Marine Sciences)
- Dr Shazrene Mohamed, South African Astronomical Observatory (Physical Sciences)
- Prof. Grant Theron, Stellenbosch University (Medical Sciences)

The P-rated researchers each also received a prize sponsored by Elsevier.

Table 1: Researchers who received NRF A-ratings in 2016

Researcher	Affiliation
A-rated for the first time	
Professor Mark Cotton	Department of Paediatrics and Child health, Stellenbosch University
Professor Erika de Wet	Institute for international and Comparative Law in Africa, University of Pretoria
Professor Bruce Hewitson	Department of Environmental and Geographical Science, University of Cape Town
Professor Florian Luca	School of Mathematics, University of the Witwatersrand
Professor Lenore Manderson	Department of Public Health, University of the Witwatersrand
Professor Bongani Mayosi	Department of Medicine, University of Cape Town
Professor Achille Mbembe	Wits Institute of Social Economic Research, University of the Witwatersrand
Professor Lynn Morris	National Institute for Communicable Diseases, National Health Laboratory Service
Professor Gerald Nurick	Department of Mechanical Engineering, University of Cape Town
Professor Craig Packer	School of Life Sciences, University of KwaZulu-Natal
Professor Chris Reason	Department of Oceanography, University of Cape Town
Professor Paul van Helden	Department of Biomedical Sciences, Stellenbosch University
Professor Willem Visser	Department of Mathematical Science, Stellenbosch University
A-rated for the second time	
Professor Nigel Bennett	Department of Zoology and Entomology, University of Pretoria
Professor David Chidester	Department of Religious Studies, University of Cape Town
Professor Pedro Crous	Department of Microbiology and Plant Pathology, University of Pretoria
Professor Michael Feast	South African Astronomical Observatory
Professor Charles Feldman	Department of Internal Medicine, University of the Witwatersrand
Professor Valerie Mizrahi	Institute of Infectious Disease and Molecular Medicine, University of Cape Town
Professor Claire Penn	School of Human and Community Development, University of the Witwatersrand
Professor Xiaohua Xia	Department of Electronic and Computer Engineering, University of Pretoria
A-rated for the third time	
Professor Leonard Barbour	Department of Chemistry and Polymer Sciences, Stellenbosch University
Professor Nicolas Beukes	Department of Geology, University of Johannesburg
Professor David Lewis-Williams	Rock Art Research Institute, University of the Witwatersrand
Professor Timothy Noakes	Department of Human Biology, University of Cape Town
Professor Helmut Prodinger	Department of Mathematical Sciences, Stellenbosch University
Professor Norman Owen-Smith	School of Animal, Plant and Environmental Sciences, University of the Witwatersrand

The Marine Protected Areas debate: Implications for the proposed Phakisa Marine Protected Areas Network

AUTHOR:
Kerry Sink[1]

AFFILIATION:
[1]South African National Biodiversity Institute, Kirstenbosch, South Africa

CORRESPONDENCE TO:
Kerry Sink

EMAIL:
k.sink@sanbi.org.za

KEYWORDS:
oceans management; fisheries; oceans economy; marine ecosystems

HOW TO CITE:
Sink K. The Marine Protected Areas debate: Implications for the proposed Phakisa Marine Protected Areas Network. S Afr J Sci. 2016;112(9/10), Art. #a0179, 4 pages. http://dx.doi.org/10.17159/sajs.2016/a0179

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

South Africa recently had the privilege of hosting prominent fisheries scientist Professor Ray Hilborn from Washington University who stimulated lively discussion on global stock status, food production, impacts of trawling on the seabed, fisheries management and Marine Protected Areas (MPAs). Professor Hilborn gave a seminar billed as 'Fisheries Myths' on 25 August 2016 and the following day participated in a formal debate at the Two Oceans Aquarium on South Africa's MPA expansion strategy and the need for additional MPAs. The debate was held between Professors Ray Hilborn and Doug Butterworth (Marine Resource and Assessment Management Group, Applied Mathematics Department, University of Cape Town) speaking against the strategy and expansion, and Dr Jean Harris (Ezemvelo KwaZulu-Natal Wildlife and Pew Fellow) and Professor Colin Attwood (Department of Biological Sciences, University of Cape Town) speaking in favour of additional MPAs. The debate examined the need for MPA expansion; the effect of MPAs on fisheries; and the role of MPAs in fisheries management, food security and biodiversity protection; and interrogated targets to increase ocean protection.

Here I examine key aspects of these discussions in the context of the proposed new Phakisa MPA Network developed through Operation Phakisa Oceans Economy – the presidential initiative to explore and unlock the economic potential of South Africa's marine and coastal environment. This network of 22 proposed MPAs was published for public comment in February this year.

Oceans under threat and the need for healthy oceans

The opening presentations of this debate by Hilborn and Harris agreed on the increasing threats to ocean ecosystems and the need for healthy oceans. The Phakisa MPA Network is centred in this common ground and aims to support the development of a sustainable oceans economy while establishing some protected ocean spaces that help create certainty and support development in the context of increasing ocean activities. It is recognised that MPAs are not a panacea and cannot address all the challenges faced by marine ecosystems.

The role of MPAs in management for healthy oceans

There are some challenges such as climate change, ocean acidification and plastic pollution that are global issues that MPAs cannot solve. MPAs are, however, important in the understanding of global change impacts and evidence is emerging that healthy ecosystems within MPAs have greater resilience to climate change.[1,2] MPAs can also reduce other global stressors such as pollution and mitigate negative effects of industrial sounds. In the proposed Phakisa MPA Network, special provisions on pollution management in the regulations will help MPAs address waste management in our oceans, but the need for wider measures is recognised and being advanced through other Phakisa initiatives.

Harris emphasised that MPAs are the only viable means to maintain intact ecosystems and undisturbed seabed communities and to protect fragile habitats from threats such as demersal fisheries and seabed mining. The important reference value of MPAs was also noted, in that research and monitoring of undisturbed areas can help us better understand impacts outside of protected areas. Attwood highlighted the importance of MPAs in providing insurance against scientific uncertainty and management failure, citing the recent discovery of a new whale species as an example of the magnitude of things we may still not know about our oceans. This is relevant for South Africa as many habitats occurring below a depth of 100 m have not been studied. New deep discoveries include fragile deep cold water coral reefs and chemosynthetic seep communities. The role of these unstudied ecosystems in the food web and life history of commercially fished species in South Africa is unknown, but the presence of cold water coral mounds in the kingklip spawning grounds may not be a coincidence. The proposed Phakisa MPA Network provides for both of these sage measures (benchmarking and precaution) in that it aims to include a proportion of such poorly known habitat.

Both sides of the debate recognised the value of MPAs in protection of habitats, although Hilborn noted the limited experimental demonstrations of broader ecosystem and biodiversity benefits of MPAs. There have, however, been multiple demonstrations, including some in our region[3,4], demonstrating how the exclusion of fishing from MPAs can influence ecosystem composition and functioning. The proposed Phakisa MPA Network aims to increase the representation of marine habitats in South Africa's protected area network. In its current form, it could provide the first protection to 46 of 54 habitats that are currently without protection. Many of these habitat types are affected by fishing, and better understanding of fisheries impacts, particularly in sensitive offshore areas, is needed. MPAs can facilitate such research.

The role of MPAs in fisheries management

The value of well-managed MPAs as a means of protecting entire ecosystems and supporting resource recovery within their bounds is established, and was agreed on by all debaters. However, the role of MPAs as a fishery management tool was one of the questions at the core of the debate. This issue has been frequently examined in the context of international policy and scientific considerations.[5-8] Unsurprisingly, most agree that MPAs will not outperform effective fisheries management in the management of fisheries resources. As Hilborn noted, MPAs are most likely to benefit fisheries where resources are seriously overfished and when the scale of fish movement is matched by the size of MPAs. In South Africa, the most depleted fisheries resources include linefish, abalone and west coast rock lobster[9] – the types of species for which some fisheries benefits through spillover from unfished

areas have been demonstrated[10,11]. Hilborn himself has co-authored several papers that show the value of spatial management in resource recovery for abalone[12], rock lobster[10] and severely overexploited species[5] and in the maintenance of spawning aggregations[5,13].

The value of MPAs in the maintenance of genetic diversity of fished species, highlighted by Attwood, was unchallenged in the debate and is recognised in international fisheries management guidelines. Maintaining genetic variability in a world with increased rates and scales of global changes seems particularly pertinent.

It is worth noting that the proposed Phakisa MPA Network is not primarily aimed at fisheries management, but is aimed at holistic environmental sustainability and mitigating the impacts of accelerated industrial activities on ocean health. The few fisheries support objectives that are included for some of the proposed MPAs, are only for species or life-history phases (e.g. the protection of spawning aggregations) for which such benefits have clearly been demonstrated. In addition, the network includes research or experimental objectives, but these did not drive selection of areas and simply take advantage of the opportunities provided by the contrast within zoned MPAs. Butterworth criticised the proposed Phakisa MPA Network for largely ignoring a key crisis in South Africa's fisheries: the poor status of the west coast rock lobster. Whilst the proposed Robben Island and Namaqua National Park MPAs do help address this issue, he is largely correct and this species may benefit from additional spatial management in the future. It should not be forgotten that existing coastal MPAs may also contribute to rock lobster recovery but improved enforcement and resolution of compliance challenges need to be addressed for current and new MPAs to contribute to the recovery of this valuable resource.

One of the criticisms of MPAs in a fisheries management context is that MPAs simply displace fishing effort which can counter fisheries management goals or displace effort into more sensitive areas.[5] The research and planning efforts for the proposed Phakisa MPA Network took cognisance of these issues and included a substantial amount of fisheries data in planning (among other sector data) to minimise effort displacement and ensure that effort is not displaced into more sensitive areas. The use of a systematic planning approach ensures spatial efficiency and the optimisation algorithms specifically select unfished and lightly fished areas for protection. They also take into account the interests of other sectors such as petroleum, mining, shipping, defence and waste management. This strategy not only reduces the impact of protection on fisheries and other stakeholders but also focuses protection into areas of good ecosystem condition; a win–win scenario. A valid criticism of any management measure applicable to both fisheries management, MPA management and any ocean management measure is non-compliance and poor governance. These challenges, often more pertinent in developing countries, require cooperative management and integrated enforcement efforts to improve compliance. Operation Phakisa includes an Integrated Enforcement Initiative that aims to support multiple compliance efforts including fisheries management and the proposed Phakisa MPA Network. Attwood reflected on the increased importance of MPAs in areas with higher biodiversity, multi-species fisheries and weaker ocean governance, making a case that no-take MPAs assume greater importance under such circumstances, particularly in developing countries.

Hilborn's recent article published in *Nature*[14] advocates that fisheries and biodiversity management should be overseen by the same body. He notes:

> *Another way to foster collaboration on a national scale would be to merge the various government departments responsible for conservation and fisheries management into a single department of marine management.*

South Africa previously had such an institution in the form of the Marine and Coastal Management (MCM) branch of the then Department of Environmental Affairs and Tourism but the mandates of this branch were split in 2010 between the Department of Agriculture, Forestry and Fisheries and the Department of Environmental Affairs. It was MCM that gave rise to the Offshore Marine Protected Area Project, which underlies much of the proposed Phakisa MPA Network. MCM also led the establishment of coastal MPAs, building on the establishment of MPAs by the former Sea Fisheries Research Institute. These steps forward in ocean protection lend support to Hilborn's view that effective biodiversity and fisheries management benefits from active collaboration between fisheries and biodiversity in area-based management.

MPAs and the South African hake trawl fishery

Hillborn reflected on the localised impact of demersal trawling in his seminar the day prior to the debate. The audience raised concerns about specific seabed habitats that are sensitive to trawling impacts and the need to better understand variability in seabed ecosystems. Hilborn agreed that sensitive ecosystems require protection and noted that 'effort should go into identifying and protecting these habitats'. The proposed Phakisa MPA Network aims to achieve just this, as emphasised by Harris in the debate. However, MPAs are not the only tools that can be used to protect habitats, and some countries have established benthic or seabed protection areas to protect vulnerable marine ecosystems *as a part* of fisheries management. South Africa still needs to develop habitat management objectives for its fisheries even though good progress has been made in understanding the potential impacts of hake trawling in South Africa. Multiple media reports covering the debate note that South Africa's deep-sea trawl industry is protecting sensitive deep-sea habitats and reducing by-catch. The proposed Phakisa MPA Network aims to provide this first protection to priority habitats affected by hake trawling in South Africa, including those habitats that are currently entirely within the trawl footprint and sensitive habitats such as cold water coral reefs. By-catch management is slowly advancing through other measures although analyses undertaken as part of the technical work for the Phakisa MPAs demonstrate that some sites in the proposed network can contribute to by-catch management and would also represent fish communities that are currently not represented in existing MPAs.[15]

MPAs and integrated oceans management

As with fisheries management, MPAs are unlikely to outperform or replace the need for other sector-specific management approaches (covering one aspect of ocean use), but they can greatly contribute to integrated ocean management. One of the sectors that was not considered by the opponents of MPA expansion was the marine mining sector. As Harris noted, MPAs, particularly the proposed Phakisa MPA Network, will help to safeguard the range of ecosystems that are subject to current and future marine mining and petroleum activities in South Africa. During audience engagement at the seminar, Hilborn noted that, in his experience, such activities operate over remarkably small areas (tens of square kilometres), but this conflicts with our experience in South Africa. When the proposed Phakisa MPA Network was developed, more than 90% of South Africa's mainland marine territory was under petroleum exploration lease. Diamond mining leases covered most of the west coast shelf and large new leases to prospect for phosphates and other minerals have been issued. The most recent phosphate lease covers more than 47 000 km^2 alone and overlaps substantially with prime hake fishing grounds. Debaters from both teams noted that MPAs are a blunter tool than more specific management actions covering a single objective. However, MPAs are very cost effective in their ability to address multiple objectives in a spatially efficient manner. Furthermore, innovation in MPA design and management can sharpen MPA effectiveness. Horizontal and vertical zonation and informed adaptive management can increase flexibility and the achievement of multiple goals. In the proposed Phakisa MPA Network, 18 of the 22 proposed MPAs are zoned with proposed zonation determined through consideration of the compatibility of particular activities with specific objectives of each MPA.

The role of MPAs in the economy

Harris noted the role of MPAs in the tourism economy, education and training, and conflict resolution. International case studies reflect on other such MPA benefits, beyond the biodiversity and fisheries aspects including their value in increasing tourism revenue, job creation, community upliftment and the preservation of culture and history.[16-18]

MPAs provide opportunities for marine ecotourism including snorkelling, scuba diving, bird and marine mammal watching, and shark tourism. Key proposed Phakisa MPAs that aim to increase benefits through their marine tourism assets and their role in the preservation of South African culture and heritage include Namaqua National Park, Robben Island, Protea Banks and Aliwal Shoal. Approaches to increase such benefits is an area of active research in South Africa and is likely to inform future protection priorities.

MPA targets

The debate chair, Kevin Cochrane (Rhodes University), framed this debate in light of South Africa's National Protected Area Expansion Strategy, a policy document published by the South African government in 2008. This strategy sets out targets for lengths of coastline and areas of ocean to receive protection, based on a 20-year protection target of 20%. Attwood shed light on the origin of this target in a fisheries context and Harris cited advances in research on targets for marine protection where, even though a range of targets has been published, all targets indicate the need for greater protection than we currently provide (see O'Leary et al.[19] for examples). Butterworth questioned the need for targets and the science behind protected area targets in particular. It was clear that there is confusion around ecosystem specific biodiversity targets, time-bound protected area targets and international action targets (such as the 10% ocean protection target set by the Convention on Biological Diversity, or the 20–30% target advocated at the Worlds Park Congress). This lack of agreement on targets notwithstanding, during the debate no-one advocated that we should have zero protection in the ocean. The proposed Phakisa MPA Network aims to advance the current 0.4% protection of South Africa's mainland ocean territory to a modest 5%. The Convention on Biological Diversity target for ocean protection is twice this amount and ecosystem-specific biodiversity targets suggest that a higher percentage is needed.[20-23] In my personal view, protracted debate around targets can detract from the real work that needs to happen to maintain ocean ecosystem health and secure the valuable ecosystem services provided by well-managed fisheries. The proportion of protected area required to achieve these goals will vary between different types of ecosystems, different species and fisheries, and will depend on the state of ecosystems and management outside of MPAs as well as many other factors. Science should continue to advance in the area of targets, without holding up implementation of protection.

Final reflections

The Phakisa MPA Network is a unique initiative developed in a unique context, with participation from 17 ministries as part of the Operation Phakisa Oceans Economy Lab. The points debated during Professor Hilborn's visit were considered by the team that planned the proposed MPA network, building on decade-long efforts to develop the more than 800 map layers that underpin the 22 proposed MPAs. The proposed Phakisa MPA Network is not designed to address overfishing. Rather, the network represents a step forward in integrated ocean management through representation of more of South Africa's diverse marine ecosystems, in areas where the last remnants of threatened ecosystems are still in good condition *and* where there is the least impact on the activities of all other stakeholders who use the ocean. The proposed Phakisa MPA Network has multiple objectives, with specific objectives set out for each individual MPA, and is aligned with fisheries management. The detailed spatial information that was used to develop the proposed network can support further work towards Marine Spatial Planning in South Africa (another Operation Phakisa initiative). The 5% protection target set by the Marine Protection Services and Ocean Governance component of Operation Phakisa is a measure to gauge progress in protected area expansion, which was recognised as an important need in light of the accelerated industrial development promoted by Operation Phakisa.

Robust debate about ocean protection targets is important, but uncertainty about the upper limits of protection targets should not delay implementation of critical protection as South Africa speeds up ocean development to ensure greater benefits from our ocean economy. Such protection must be designed in collaboration not only with fisheries but also with other ocean-use sectors. The proposed Phakisa MPA Network seeks to achieve this through integrated planning covering all affected marine sectors and through optimisation algorithms that minimise impacts of MPAs on all stakeholders, to deliver a spatially efficient MPA network that achieves multiple protection objectives. The opportunity to expand South Africa's MPA network exists, the need is urgent, and the proposed Phakisa MPA Network stands to provide well-focused protection based on sound evidence and a strongly participatory process.

References

1. Micheli F, Saenz-Arroyo A, Greenly A, Vazquez L, Montes JAE, Rosetto M, et al. PLoS ONE. 2012;7(7), e40832, 8 pages. http://dx.doi.org/10.1371/journal.pone.0040832
2. Mellin C. Marine protected areas increase resilience among coral reef communities. Ecol Lett. 2016;19(6):629–637. http://dx.doi.org/10.1111/ele.12598
3. Lasiak TA, Field JG. Community-level attributes of exploited and unexploited rocky infratidal macrofaunal assemblages in Transkei. J Exp Mar Biol Ecol. 1995;185:33–53. http://dx.doi.org/10.1016/0022-0981(94)00130-6
4. Götz A, Kerwath SE, Attwood CG, Sauer WH. Effects of fishing on a temperate reef community in South Africa 2: Benthic invertebrates and algae. Afr J Marine Sci. 2009;31:253–262. http://dx.doi.org/10.2989/AJMS.2009.31.2.13.885
5. Hilborn R, Stokes K, Maguire JJ, Smith T, Botsford LW, Mangel M, et al. When can marine reserves improve fisheries management? Ocean Coast Manage. 2004;47(3):197–205. http://dx.doi.org/10.1016/j.ocecoaman.2004.04.001
6. Food and Agriculture Organization of the United Nations (FAO). Fisheries management: Marine protected areas and fisheries. FAO Technical Guidelines for Responsible Fisheries No. 4, Suppl. 4. Rome: FAO; 2011. Available from: http://www.fao.org/docrep/015/i2090e/i2090e.pdf
7. Rice J, Moksness E, Attwood C, Brown SK, Dahle G, Gjerde KM, et al. The role of MPAs in reconciling fisheries management with conservation of biological diversity. Ocean Coast Manage. 2012;69:217–230. http://dx.doi.org/10.1016/j.ocecoaman.2012.08.001
8. Caveen A, Polunin N, Gray T, Stead SM. The controversy over marine protected areas: Science meets policy. Berlin: Springer; 2014.
9. Department of Agriculture, Forestry and Fisheries. Status of stocks report [document on the Internet]. c2014 [cited 2016 Sep 12]. Available from: http://www.nda.agric.za/doaDev/sideMenu/fisheries/03_areasofwork/pdf
10. Goni R, Hilborn R, Díaz D, Mallol S, Adlerstein S. Net contribution of spillover from a marine reserve to fishery catches. Mar Ecol Prog Ser. 2010;400:233–243. http://dx.doi.org/10.3354/meps08419
11. Kerwath SE, Winker H, Götz A, Attwood CG. Marine protected area improves yield without disadvantaging fishers. Nat Commun. 2013;4:2347–3347. http://dx.doi.org/10.1038/ncomms3347
12. Hilborn R, Orensanz JL, Parma AM. Institutions, incentives and the future of fisheries. Philos Trans R Soc Lond B. 2005;360(1453):47–57. http://dx.doi.org/10.1098/rstb.2004.1569
13. Hilborn R. Moving to sustainability by learning from successful fisheries. AMBIO. 2007;36(4):296–303. http://dx.doi.org/10.1579/0044-7447(2007)36[296:MTSBLF]2.0.CO;2
14. Hilborn R. Marine biodiversity needs more than protection. Nature. 2016;535(7611):224–226. http://dx.doi.org/10.1038/535224a
15. Solano-Fernandez S, Attwood CG, Chalmers R, Clark BM, Cowley PD, Fairweather T, et al. Assessment of the effectiveness of South Africa's marine protected areas at representing ichthyofaunal communities. Environ Conserv. 2012;39(3):259–270. http://dx.doi.org/10.1017/S0376892912000070
16. Leisher C, Van Beukering P, Scherl LM. Nature's investment bank: How marine protected areas contribute to poverty reduction [document on the Internet]. c2007 [cited 2016 Sep 12]. Available from: http://www.nature.org/media/science/mpa_report.pdf
17. McCook LJ, Ayling T, Cappo M, Choat JH, Evans RD, De Freitas DM, et al. Adaptive management of the Great Barrier Reef: A globally significant demonstration of the benefits of networks of marine reserves. Proc Natl Acad Sci USA. 2010;107(43):18278–18285. http://dx.doi.org/10.1073/pnas.0909335107

18. Costello MJ. Long live marine reserves: A review of experiences and benefits. Biol Conserv. 2014;176:289–296. http://dx.doi.org/10.1016/j.biocon.2014.04.023

19. O'Leary BC, Winther-Janson M, Bainbridge JM, Aitken J, Hawkins JP, Roberts CM. Effective coverage targets for ocean protection. Conserv Lett. 2016, Art. #12247. http://dx.doi.org/10.1111/conl.12247

20. Rondinini C, Chiozza F. Quantitative methods for defining percentage area targets for habitat types in conservation planning. Biol Conserv. 2010;143:1646–1653. http://dx.doi.org/10.1016/j.biocon.2010.03.037

21. Harris L, Nel R, Holness S, Sink K, Schoeman D. Setting conservation targets for sandy beach ecosystems. Estuar Coast Shelf Sci. 2014;150:45–57. http://dx.doi.org/10.1016/j.ecss.2013.05.016

22. Metcalfe K, Delavenne J, Garcia C, Foveau A, Dauvin J, Coggan R, et al. Impacts of data quality on the setting of conservation planning targets using the species–area relationship. Divers Distrib. 2013;19:1–13. http://dx.doi.org/10.1111/j.1472-4642.2012.00921.x

23. Karenyi N, Nel R, Altwegg R, Sink K. Incorporating species detectability into conservation targets based on the species–area relationship. Divers Distrib. 2016;22(7):758–769. http://dx.doi.org/10.1111/ddi.12441

AUFWIND: An ambitious German microalgae project for producing third-generation biofuels

AUTHOR:
Johan U. Grobbelaar[1]

AFFILIATION:
[1]Department of Plant Sciences, University of the Free State, Bloemfontein, South Africa

CORRESPONDENCE TO:
Johan Grobbelaar

EMAIL:
grobbeju@ufs.ac.za

KEYWORDS:
biotechnology; photobioreactor; *Chlorella*; lipids

HOW TO CITE:
Grobbelaar JU. AUFWIND: An ambitious German microalgae project for producing third-generation biofuels. S Afr J Sci. 2016;112(9/10), Art. #a0174, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0174

The term 'microalgae' has no taxonomic meaning, except that it refers to a spectrum of photosynthesising micro-organisms from Protista and Archaebacteria, including chlorophyta (green algae) and cyanobacteria (blue-green algae). Microalgae have been studied in the laboratory, in mass outdoor cultures and in nature for more than a century. They are grown for a variety of potential applications, which include basic research on photosynthesis, the production of lipids for energy, bioremediation, anti-microbial substances, cheap protein for animal and human nutrition, and the production of various bio-chemicals.[1,2]

Microalgae have several competitive advantages over higher plants, namely higher growth rates, suitability for growing in photobioreactors, and non-reliance on good agriculture soils. Storage products of microalgae generically could be either carbohydrates and/or lipids, with the potential to produce ethanol, diesel, methane and even kerosene from these compounds. However, a major frustration for microalgal biotechnologists has been the realisation of much lower yields than what laboratory measurements suggest should be possible.[3]

The potential of algae as a fuel source is undisputed. It was these photo-autotrophic micro-organisms that were fixing sunlight energy into lipids for millions of years, generating the petroleum reserves that modern human civilisation uses today. However, such reserves are finite, and the challenge is to marry biology with technology to produce economically competitive fuels without harming the environment or compromising our food security. The fundamental ability of microalgae to produce energy-rich biomass from carbon dioxide, nutrients and sunlight, through photosynthesis for biofuels, has led to the concept of 'third-generation biofuels'.

The key compounds used for bio-diesel and kerosene production are lipids, especially the triacylglycerols (referred to as TAGs). These lipids, once extracted, need to be trans-esterified for biodiesel, and a further 'cracking' step is required to produce kerosene. For biofuels, microalgae with high TAG content are required. A number of such algae have been isolated and lipid contents from 20% to 60% have been achieved. An essential step in forcing microalgae to channel energy compounds into lipids is some form of stress. Stressors such as high light, low nitrogen, low phosphorus and high salinity have been used. Limiting nitrogen was the main stressor in the cultures used in our study. This limitation was achieved by transferring exponentially growing cells into a balanced nutrient medium with a known limiting concentration of nitrogen.

The AUFWIND project at the Research Centre Jülich in Jülich, Germany consists of three different commercially available photobioreactor types, constructed adjacent to each other. This enabled us compare three very different pilot-to-commercial-scale systems for biomass and lipid production. The project is undisputedly the first of its kind to address biofuel production from microalgae on such a large scale and in such different systems.

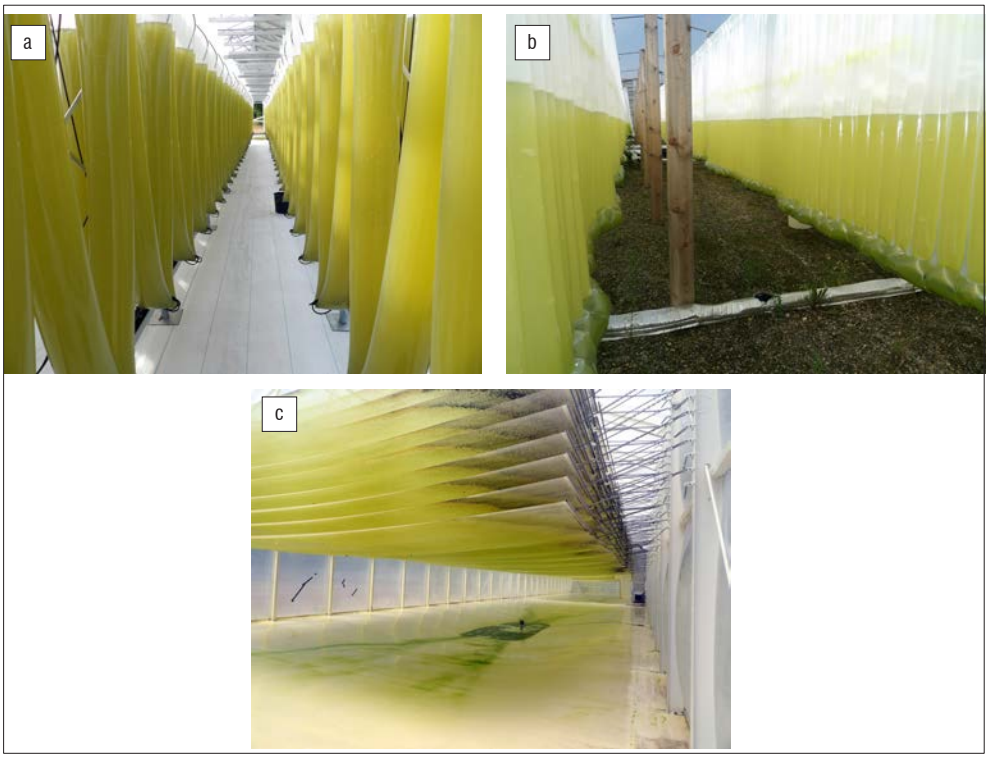


Figure 1: Three photobioreactors at the AUFWIND project. (a) Novagreen photobioreactor, housed in a glasshouse, comprising 1338 V-bags each containing 25 L culture (note the yellow colour, typical of lipid-rich microalgae); (b) the Phytolutions photobioreactor is made up of six sections, each containing up to 10 'curtains' of vertical tubes interconnected with a horizontal tube at the bottom; and (c) stacked horizontal mesh sheets of the IGV system, housed in a plastic-covered greenhouse.

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

The three photobioreactors are from Novagreen[4] (Figure 1a), Phytolutions GmbH[5] (Figure 1b), and Institüt für Getreideverarbeitung GmbH (IGV)[6] (Figure 1c). Each photobioreactor occupies 500 m^2 of land surface area. The Novagreen system consists of interconnected vertical plastic tubes, each roughly 150 mm in diameter, whereas the Phytolutions system is outdoors and consists of 'curtains' of vertical plastic tubes, each with a diameter of about 90 mm. The most ambitious photobioreactor is from IGV, and consists of horizontally layered nets; the algae are sprayed over the nets and allowed to grow while dripping from one net to the next. All systems received additional carbon dioxide, and an array of environmental and culture parameters were measured continuously.

The green alga *Chlorella vulgaris* (Beijerinck), strain CCALA 256, was used as a growth organism. One of the main tasks was to manipulate growth conditions so that the microalgae converted their stored energy into lipids, and to establish protocols to run the various photobioreactors on a large scale. This was accomplished in just over 2 months of intensive experimentation that resulted in modifications to the designs of the photobioreactors, different microalgal strain selection, and the replacement of the nutrient broth with a so-called balanced one (so that at final yield all the supplied nutrients had been taken up by algae). From the experiments it became evident that the specific growth rate of the algae was constrained by an upper limit that differed for each photobioreactor, but was always approximately linearly proportional to the daily dose of photosynthetically active radiation. Extremely encouraging was the fact that the microalgae could be manipulated on an industrial scale and lipid contents >40% were achieved.

One should have no illusions regarding the technology and economic feasibility of the project. However, with continued research, optimisation, and utilisation of waste resources, it is highly likely that the production of lipids from microalgae for biofuels can become a reality in the near future. An added benefit from the project is the generation of valuable data for large-scale industrial applications.

References

1. Grobbelaar JU. Microalgal biomass production: Challenges and realities. Photosynth Res. 2010;106:135–144. http://dx.doi.org/10.1007/s11120-010-9573-5
2. Grobbelaar JU. Microalgae mass culture: The constraints of scaling up. J Appl Phycol. 2012;24:315–318. http://dx.doi.org/10.1007/s10811-011-9728-6
3. Richmond A. Efficient utilization of high irradiance for production of photoautotrophic cell mass: A survey. J Appl Phycol. 1996;8:381–387. http://dx.doi.org/10.1007/BF02178581
4. Novagreen Projektmanagement GmbH. Industrial production of microalgae [homepage on the Internet]. No date [cited 2016 Aug 22]. Available from: http://www.novagreen-microalgae.de
5. Phytolutions [homepage on the Internet]. No date [cited 2016 Aug 22]. Available from: http://www.phytolutions.com
6. Institut für Getreideverarbeitung GmbH [homepage on the Internet]. No date [cited 2016 Aug 22]. Available from: http://www.igv-gmbh.de

The Piltdown case: Further questions

AUTHOR:
J. Francis Thackeray[1]

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

CORRESPONDENCE TO:
J. Francis Thackeray

EMAIL:
Francis.Thackeray@wits.ac.za

KEYWORDS:
Piltdown Man; forensic analyses; palaeontology; Sussex; Teilhard de Chardin

HOW TO CITE:
Thackeray JF. The Piltdown case: Further questions. S Afr J Sci. 2016;112(9/10), Art. #a0170, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0170

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

About 60 years ago, a South African anatomist, Joseph Weiner, published a book entitled *The Piltdown Forgery*[1], exposing a hoax that had been perpetrated about 100 years ago at the site of Piltdown in Sussex, England. The announcement of 'Piltdown Man' – classified as *Eoanthropus dawsoni* and believed to be a hominid apparently associated with Pleistocene fauna – had been made by Smith Woodward of the British Museum (Natural History) at Burlington House in London on 18 December 1912. However, it turned out that the 'hominid' was a fabrication in which a subfossil human cranium and a modern orangutan jaw (both stained brown) were placed together in a gravel pit, thereby confusing palaeontologists. In the process, Raymond Dart's announcement of the 'Taung Child' (*Australopithecus africanus*) from South Africa was disregarded by many (including the anatomist Sir Arthur Keith) who questioned Dart's claim that this small-brained fossil represented a genuine Pleistocene hominid.

Weiner, together with his colleagues Kenneth Oakley (Keeper of Palaeontology at the British Museum) and Sir Wilfred Le Gros Clark, recognised the forgery. After investigation, they concluded that Charles Dawson (a lawyer and amateur archaeologist based in Sussex) was the prime suspect responsible for the forgery. This conclusion has been endorsed recently by Isabelle de Groote and her palaeoanthropologist colleagues[2] who have undertaken high-tech forensic analyses, including the use of micro-CT scanning of teeth and DNA analyses of the craniodental material. They conclude that only a single hoaxer was responsible, namely Dawson, now known to have been responsible for more than 30 forgeries, perhaps in the hope of earning him the distinction of being elected as a Fellow of the Royal Society (which he never attained).

In the light of the recent forensic analyses, one may well accept a verdict that Dawson was indeed the sole perpetrator: once and for all, Piltdown case closed.

But is it closed? Several people other than Dawson have been questioned, because they were perhaps directly implicated or knew about the forgery. As shown by Thackeray[3], Robert Essex, Louis Leakey, Stephen Jay Gould and Thackeray himself have all been suspicious of Pierre Teilhard de Chardin, a French Jesuit palaeontologist who was known to have been a 'joker', and who was based at Ore Place, a theological seminary near Hastings in Sussex, within 50 km of Piltdown where Teilhard de Chardin contributed to excavations in 1912 (when he found a fossil elephant molar) and again in 1913 when he found the orangutan canine (assumed to be a hominid associated with the human cranium).

Some details are pertinent here. In the first week of January 1913, Teilhard de Chardin wrote an essay in *Etudes* beginning with the words 'There was a time when *La Prehistoire* deserved to be suspect, and the subject of jokes'. His entire essay is about the current understanding of human evolution, but most strangely (and suspiciously) he omits all reference to Piltdown even though it had been officially announced to great acclaim just 3 weeks earlier. Almost immediately after that announcement, Teilhard de Chardin wrote to his friend Felix Pelletier (with whom he had collected fossils in Sussex), saying: 'We must do nothing. We must wait for the criticisms that will follow. Marcellin Boule [an eminent French prehistorian] will not be taken in.' This essentially convinced Thackeray[3] that Teilhard de Chardin knew from the very beginning that 'Piltdown Man' was not genuine.

In 1977, Oakley told Thackeray that he was giving Teilhard de Chardin the benefit of the doubt because he was a priest. But Teilhard de Chardin was a Jesuit, whom Phillip Tobias called a 'joker', and Jesuits were allegedly allowed to lie providing it was a joke.[3]

Teilhard de Chardin and Martin Hinton allegedly said they *knew* who had been the Piltdown perpetrator, and both were adamant that it was *not* Dawson.[3] The question arises – were they 'in on the joke', and was this joke directed *against* Dawson, especially if some people known to Dawson began to suspect his forgeries (more than 30 of them)?

This idea is relevant to the fact that the Piltdown orangutan jaw is likely to have come from Borneo (as indicated from the recent DNA analyses).[2] Secondly, it is pertinent to facts presented by Sherratt[4] who referred to an expedition led by Everett in 1878, when material (including orangutan crania and mandibles) was collected in Borneo and brought back to England. Most if not all of this material was deposited in the British Museum (Natural History), but evidently 'duplicates' could have been distributed elsewhere, subject to the decision of a committee.[4] It would seem that such 'duplicates' could have been distributed to donors of the expedition, including members of the Willett family who lived in Sussex, notably Henry Willett (an antiquarian whose wealth was based on his successful brewery, supplemented by his successful investment in railways); Ernest Willett (who had a strong interest in ancient coins); and Edgar Willett who was trained at Oxford, practised as an anaesthetist, and also served as a curator of a museum with expertise in anatomy.

On the death of Henry Willett, his son Edgar appears to have taken early retirement and enjoyed the life of a gentleman in Sussex, pursing his pleasure in sport and croquet. And here is the rub: *he is known to have accompanied Dawson at Piltdown*.

There is strong reason to be interested in the Willett family, especially Edgar Willett whose background in anatomy is intriguing, and his association with Dawson at Piltdown is remarkable. As the Willetts (among others, including Darwin) had provided funds for the Everett expedition to Borneo, it would seem entirely possible that a 'duplicate' orangutan mandible from that expedition found its way into the family's collections.

The family name Willett is strongly associated with Ore in Sussex, and Ore Place in Ore (near Hastings in Sussex) is where Teilhard de Chardin was based between 1908 and 1912, the very years in which Piltdown material was initially collected. Work resumed in August 1913 when Teilhard de Chardin was invited to assist in excavations, and when he found the canine. The question now arises as to whether the French Jesuit knew about a Piltdown joke

through one of the Willetts, perhaps Edgar Willett who (as an anatomist) could have been in a position to facilitate a joke against Dawson, potentially with Teilhard de Chardin as an adviser.

Thackeray[3] has contended that Teilhard de Chardin may have wanted the Piltdown joke to be exposed without damaging his future career in palaeontology. When he was invited to Piltdown in August 1913, to continue excavations, he may have 'discovered' the reddish-brown canine in a deliberate attempt to expose the whole thing as a joke. However, he found the canine in an area that had already been thoroughly searched in the exposed gravels. Woodward had initially expressed disbelief that he had found the canine there, but was subsequently taken in.

Gould[5] was strongly suspicious of Teilhard de Chardin, especially because the French palaeontologist had stated in 1920 that the condyle of the Piltdown mandible might have been deliberately broken. In Gould's view, this statement was tantamount to an admission that Teilhard de Chardin knew of the forgery although such a thing was not suspected at the time.

Thackeray[3] has referred to two letters that Teilhard de Chardin wrote shortly before his death in 1955, after exposure of the hoax. In a letter to Oakley dated 28 November 1953, Teilhard de Chardin writes:

> *Would it have been impossible for some collector who had in his possession some ape bones, to have discarded specimens into the pit? The idea sounds fantastic. But, in my opinion, no more fantastic than to make Dawson the perpetrator of the hoax.*

Similarly, on 8 December 1953, Teilhard de Chardin wrote to Abbe Henri Breuil, saying:

> *I have difficulty in accepting a hoax [supercherie] on the part of Dawson (a close friend of Smith Woodward). And as fantastic as it seems, I admitted rather that someone threw, innocently, from the cottage nearby, some 'collection' in the 'Pit' of Piltdown.*

The question arises: was Teilhard de Chardin referring to an actual 'collection' of material that had been in the hands of 'some collector' including a member of the Willett family – perhaps Edgar Willett, an anatomist and a man of leisure in Sussex, who knew Dawson?

Thackeray[3] refers to a letter that was allegedly written by Teilhard de Chardin and deposited in a bank, with instructions that it be opened only after his death. No such letter was announced after his death. He died in New York when he was based there, at the Wenner Gren Foundation. Leslie Aiello (current President of the Foundation) told Thackeray (written communication, 2016 September 07) that Teilhard de Chardin's documents were removed from his office by Jesuits after his death, and bank details are not available. Please could someone try to find out where Teilhard de Chardin did his banking transactions before his death in New York? This might truly and finally bring the Piltdown case to a close.

References

1. Weiner JS. The Piltdown forgery. London: Oxford University Press; 1955.
2. De Groote I, Girdland Flink L, Abbas R, Bello SM, Burgia L, Buck LT, et al. New genetic and morphological evidence suggests a single hoaxer created 'Piltdown man'. R Soc Open Sci. 2016;3, Art. #160328, 14 pages. http://dx.doi.org/10.1098/rsos.160328
3. Thackeray JF. Deceiver, joker or innocent? Teilhard de Chardin and Piltdown Man. Antiquity. 2012;86:228–234. http://dx.doi.org/10.1017/S0003598X0006258X
4. Sherratt A. Darwin among the archaeologists: The John Evans nexus and the Borneo Caves. Antiquity. 2002;76(291):151–157. http://dx.doi.org/10.1017/S0003598X00089924
5. Gould SJ. The Piltdown conspiracy. Natural History Magazine. 1980;89:8–28.

ASSAf 1996–2016
Commemorative article

Finding an influential voice for academies in Africa

AUTHOR:
Robin M. Crewe[1]

AFFILIATION:
[1]Centre for the Advancement of Scholarship, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Robin Crewe

EMAIL:
Robin.Crewe@up.ac.za

KEYWORDS:
ASSAf; IAP; NASAC; advisory role; evidence-informed policy

HOW TO CITE:
Crewe RM. Finding an influential voice for academies in Africa. S Afr J Sci. 2016;112(9/10), Art. #a0175, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0175

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

Scholarly academies have been in existence for about 350 years, with the oldest being those that were established in Europe in the 17th century. These institutions consist of groups of individuals who are elected by their peers to be members (often called fellows); since the middle of the 19th century, election to the august ranks of most academies has been based on recognition of the outstanding scholarly work done by those proposed for membership. Academies have made the transition from being learned societies to being select groups of eminent scholars who are often widely admired in their countries.

The transformation of academies into institutions that have the most eminent scholars of a nation as their members, meant that they became influential in advising their national governments and were regarded as a unique source of expert advice in matters of national concern. However, this has not always been the case in the history of European academies and often is not the case for more recently established academies in other parts of the world, particularly in Africa. Lorna Casselton, a former Foreign Secretary of the Royal Society of London pointed out somewhat acerbically, 'My post was instituted in 1723, nearly 60 years before the British government appointed its first Secretary of State for Foreign Affairs', making the point that academies were transnational in their reach long before governments formally established offices to deal with international relations. Academies engage with a global community of scholars whose interests are aligned primarily with their disciplines and only secondarily with their national origins. Governments have often been slow in understanding the value that academies have in providing advice. A good example of this is the publication by the Academy of Science of South Africa (ASSAf) of the report *HIV/AIDS, TB and Nutrition* (2007) which, although not accepted initially, had a significant impact through providing a basis for a radical change in national policy for the treatment of HIV/AIDS.

The use of independent academies to provide advice on matters of public interest is of significant benefit to governments when they are confronting difficult political and technological choices. However, the degree to which this advice is solicited and then subsequently used depends on the nature of the government in place and their assessment of the political risks associated with accepting independent, impartial advice.

With the exception of the US National Academies of Science (USNAS) that was established by Abraham Lincoln to offer advice to Congress on matters of science and technology, most academies started out their lives as learned societies that were largely honorific in their function. The rigour with which their members are chosen, meant that there was significant prestige associated with election to these bodies. However, academies did not conceive of themselves as reviewing matters of national importance based on the best available evidence with a view to providing informed insights for policy development by governments. This latter role has developed and accelerated over the course of the 20th century, as technological advances and growth of human populations have placed increasing strain on natural resources, thereby requiring governments to make policy choices in relation to highly technical matters. The range of expertise that academies can bring to bear on the assessment of these problems through their membership is unique and valuable, if appropriately used.

The key element that makes academies valuable as sources of multi-perspective advice is their independence in two important respects – they are governed by councils that are elected by their members and they have a professional secretariat appointed by their councils.

As a subset of the global family of academies has re-conceptualised their role as outlined above, they have realised the necessity to establish networks of academies in order to address issues that are of regional or global concern. The InterAcademy Panel (now IAP for Science) that is the global network of science academies was established in 1993 in order to assist with the coordination of the activities of national academies on a global scale. It now has 107 member academies and plays an important coordinating role in matters that are of global concern by convening meetings of its member academies and by facilitating the establishment of regional academy networks to deal with specific matters at a regional level.

In the case of Africa, although some academies have been in existence for over a century, most have a much shorter history associated with the timing of their country's liberation from their colonial governments. Conceptualising a role beyond the honorific one by these academies has been a recent phenomenon that is still in the process of being formally established.

A successful advisory role for academies is dependent on three conditions: they need to be seen as offering independent advice which is not partisan, they need to have well-established methodologies for providing advice that is robust and establishes confidence in the reports that are produced and, finally, they need to have a government and civil society that is potentially receptive to the advice.

Based on this short exploration of the history of academies, I can now turn to ASSAf – the Academy of Science of South Africa – and put its development in the context that I have sketched. ASSAf was inaugurated in 1996 at a gala dinner hosted by then President Nelson Mandela who acted as the patron of this newly established academy. The key point about the launch of ASSAf was that it conceptualised itself *ab initio* in a way that was not common for academies – it had the traditional role of honouring those who were elected to its membership, but it also defined for itself an activist role of using science for the benefit of society.

Between 1996 and 2001, the Members of ASSAf were in discussion with the officials of the Department of Science and Technology (DST), actually the Department of Arts, Culture, Science and Technology at the time, to get an Act passed by parliament to establish ASSAf as a statutory body. This was finally done in 2001 and ASSAf came into being as the national academy of science in May 2002. During this period, ASSAf was largely pursuing the honorific

role that was the key element that gives academies their strength and their substance – the expertise and standing of their members.

With the establishment of ASSAf as a statutory body, it entered a new phase of development during which a professional secretariat was established and an executive officer appointed. The first Executive Officer, Prof. Wieland Gevers, was instrumental in giving ASSAf an institutional identity through the writing of the regulations that govern its activities and in giving substance to the work of the secretariat. Indeed, he translated the activist aspirations of this nascent academy into action through the initiation of the first study that ASSAf undertook: *Report on a Strategic Approach to Research Publishing in South Africa*. This report was commissioned by DST, and when completed was welcomed both by that department and the Department of Higher Education and Training (DHET) because its recommendations provided a basis for quality assurance of journals published in South Africa and provided DHET with recommendations for a more reliable basis for evaluating the research publications produced by universities.

ASSAf's ability to have an impact in influencing government policy was given a very significant boost through being one of the academies included in the African Science Academies Development Initiative (ASADI) that was initiated by USNAS with funding from the Bill & Melinda Gates Foundation. The programme provided funding for 5 years and also involved intensive mentoring of the staff of the ASSAf secretariat by the staff of USNAS. During this period, the way in which ASSAf provided advice to government and other organisations, evolved to the point at which a range of instruments was deployed to provide advice in different situations.

The other important element of the ASADI programme was that its annual meetings provided a platform for the expansion of the Network of African Academies of Science (NASAC) that was established in 2001 with ASSAf as one of the founding nine members. At the conclusion of the ASADI programme in 2015, NASAC had grown to include 21 African academies of science and had instituted a range of programmes that fostered both academy development and the collaboration of member academies in a range of studies including addressing the issues of the use of scarce water resources and maternal and child health.

The real challenge facing NASAC and its member academies is the way in which they will be able to influence governments in Africa and the pan-African organisation, the African Union (AU), to address the Sustainable Development Goals and achieve the AU Commission's *Agenda 2063* aspirations. Up to this point, their role in these discussions has not been central, but they will need to conceptualise ways in which they will be able to achieve a much greater level of influence over the coming decade because ambitious targets – 'no poverty' and 'zero hunger' – have already been set for 2030.

Working in their own national environments and collectively as members of NASAC (the affiliate network of the IAP for Africa), the African academies of science will need to generate a set of credible interventions and recommendations that will assist governments and the AU to achieve their goals. Goal setting of this kind can be depressing if the goals are not achieved, but the achievement of the goals would lead to a level of well-being in the countries involved that would be its own reward.

At the beginning of this piece, I posited three requirements for a modern academy to be successful: the first was that the advice should be seen to be impartial, the second was that the academy should have an armoury of instruments that could be used to generate advice and the third was a government and civil society that was receptive of advice. I believe that the first two requirements are already in place for most of the African academies of science thanks to their participation in the activities of NASAC. The third remains somewhat problematic, as this is an area in which the academies and their members need to use their influence in order to ensure that their recommendations and statements are taken seriously. This can only be achieved by an ongoing engagement with the individuals and institutions that need to be influenced – ensuring that the voice of the academies is heard not only in the national context, but also on regional and continental scales.

Response to Thackeray (2016) – The possibility of lichen growth on bones of *Homo naledi*: Were they exposed to light?

AUTHORS:
Patrick S. Randolph-Quinney[1,2,3]
Lucinda R. Backwell[2]
Lee R. Berger[2]
John Hawks[2,4]
Paul H.G.M. Dirks[2,5]
Eric M. Roberts[2,5]
Godwin Nhauro[2]
Jan Kramers[6]

AFFILIATIONS:
[1]School of Forensic and Applied Sciences, University of Central Lancashire, Preston, Lancashire, United Kingdom
[2]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa
[3]School of Anatomical Sciences, University of the Witwatersrand Medical School, Johannesburg, South Africa
[4]Department of Anthropology, University of Wisconsin-Madison, Madison, Wisconsin, USA
[5]Department of Geoscience, James Cook University, Townsville, Queensland, Australia
[6]Department of Geology, University of Johannesburg, Johannesburg, South Africa

CORRESPONDENCE TO:
Patrick Randolph-Quinney

EMAIL:
prandolph-quinney@uclan.ac.uk

KEYWORDS:
Rising Star Cave; mineral staining; taphonomy; sedimentology

HOW TO CITE:
Randolph-Quinney PS, Backwell LR, Berger LR, Hawks J, Dirks PHGM, Roberts EM, et al. Response to Thackeray (2016) – The possibility of lichen growth on bones of *Homo naledi*: Were they exposed to light? S Afr J Sci. 2016;112(9/10), Art. #a0177, 5 pages. http://dx.doi.org/10.17159/sajs.2016/a0177

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

Thackeray[1] questions the hypothesis of deliberate body disposal in the Rising Star Cave by *Homo naledi,* as proposed by Dirks and colleagues[2]. Thackeray proposes that lichens produced mineral staining on the skeletal remains of *H. naledi*. As lichens require some exposure to light, in Thackeray's opinion, the presence of mineral staining necessitates either a direct entrance deep into the Rising Star Cave that once admitted light into the Dinaledi Chamber, or relocation of mineral-stained bones from a location exposed to light. Here we consider multiple lines of evidence that reject Thackeray's hypothesis that lichens deposited mineral staining upon the surface of these skeletal remains. We welcome the opportunity to address the inferences presented by Thackeray, and further hope that this response may dispel misinterpretations of our research[2], and of other areas of the scientific literature that bear upon site formation processes at work within the Rising Star Cave system.

Briefly, we review Thackeray's[1] chain of inference: (1) lichens can be found today growing as colonies on the surface of chert, dolomite and sub-aerially exposed bones found within the Cradle of Humankind; (2) some colonies of lichen on extant rocks overlie deposits of black manganese oxy-hydroxide on the rock surfaces; (3) some manganese (Mn) mineral deposits appear as diffuse spots, which have some resemblance to the shape and surface patterning of some lichen colonies; (4) small dots or spots of Mn present on the surface of the bones of *H. naledi* may, therefore, have been produced as a consequence of lichen growth; (5) the growth of lichen colonies requires light, either in the open environment or within the light zone of caves; (6) by extension, the growth of lichen on the bones of *H. naledi* required the presence of subdued, but essential lighting; (7) the presence of such lighting indicates the existence of a second entrance into the Dinaledi Chamber; therefore (8) a re-assessment of the deliberate body disposal hypothesis is required.

We must state from the outset that we have no disagreement with conditions (1) to (3) above, which comport with our own first-hand observations from geological outcrops within the Cradle of Humankind. Lichens do indeed grow on chert, dolomite and bone; they do sometimes overlie Mn deposits; and there is often a diffuse spot pattern in their surface distribution. But available evidence is not consistent with the rest of this line of reasoning. We here combine previously published taphonomic and geological data from the Dinaledi Chamber[2], together with a fuller review of the literature and re-interpretation of photographic evidence presented by Thackeray[1].

Abundant actualistic research documents mechanisms of Mn deposition on bone that have nothing to do with lichens.[3] Mn and iron (Fe) deposition on bone in dark, wet, alkaline cave contexts like that found in the Dinaledi Chamber is normally a result of diagenetic recrystallisation of bone, incorporating trace elements from surrounding sediment and water via a diffusion–absorption process.[4] In dolomitic caves, which have relatively high pH and oxidising redox, the alkali metals sodium and potassium and the alkaline earth metals calcium and magnesium are the most abundant soluble cations, while the transition metals copper, Fe and Mn are the least abundant, and the alkaline earth metals strontium and barium, and the transition metal zinc are of intermediate availability.[5] Under these cave conditions, the stable forms of Fe and of Mn are hydrates and manganese oxide compounds, respectively. Precipitation of these highly insoluble hydrates and oxides varies with pH and tends to form crusts and coatings.[6] Mn is mainly located on the surface and crack edges of fossil bones, whereas Fe shows deeper penetration into the bone matrix.[7,8] Mn, which naturally occurs in groundwater, enters the bone environment as mobile Mn^{2+}. In the presence of free O_2, the oxidation of Mn^{2+} to Mn^{3+} and ultimately to insoluble Mn^{4+} is thermodynamically favourable, but proceeds extremely slowly unless mediated by microbial action, which is chemotrophic and does not require light.[9,10] Oxidised Mn precipitates as a number of manganese oxides and hydroxides.[11] Similarly, under the same conditions, mobile ferrous iron (Fe^{2+}) approaching the bone may be oxidised to ferric iron (Fe^{3+}) which rapidly precipitates as limonite ($FeOOH.nH_2O$), which later usually undergoes transformation to goethite (FeOOH). Alternatively, pyrite (FeS_2) is formed instead of haematite via precipitation of iron sulfide as decaying collagen releases sulfide ions into solution.[12] Black or dark brown colour of fossil bones is a result of high levels of pyrite and manganese and iron oxide and hydroxide coatings. To summarise, processes that do not involve lichens are sufficient to explain the presence of Mn stains on fossil bone in contexts like that found within the Dinaledi Chamber today.

Mn, Fe and other mineral deposition occurs on a large fraction of the skeletal material from the Dinaledi Chamber in a variety of depositional forms. Thackeray[1] chose to illustrate the pattern of Mn staining (specifically spotting) on only a single specimen. In his Figure 6 he provides a photograph of tibial specimen U.W. 101-996 (his figure caption misidentifies this specimen as U.W. 101-1070) with the figure legend:

> *Tibial shaft specimen U.W. 101-1070* [sic] H. naledi *from the Dinaledi Chamber, with dotted coatings of manganese oxy-hydroxide. It is suggested that the black dots result, at least in part, from the growth of lichen as a bacterial-algal-fungal symbiont that includes a photobiont. The growth of lichen on such bone surfaces, even for a limited time, may have occurred in subdued, but essential lighting. Note the* distribution [our emphasis] *of manganese oxy-hydroxide, extending from a continuous matt to more dotted occurrences; this pattern is* potentially analogous [our emphasis] *to the dispersal of lichen from a central thallus.*[1]

However, the distribution of Mn staining is not only on this one side of the specimen, but on all anatomical sides (anterior, posterior, medial and lateral), as shown in Figure 1. Such a distribution can also be seen on many specimens (for example, in Figure 2, specimen U.W. 101-1070, correctly attributed here). The distribution of Mn staining on the Dinaledi Chamber material is generally circumferential, occurring on multiple sides of bones. That distribution is not compatible with lichen growth in the present context, and there are significant reasons to reject the hypothesis that the Dinaledi Chamber is a secondary deposit.[2,13]

(a) Medial aspect; (b) lateral aspect; (c) close-up of anterior shaft; (d) mid shaft; (e) close-up of distal end seen in (a); (f) close-up of distal end seen in (b).

Figure 1: Patterns of mineral staining affecting tibia U.W. 101-996. Note the distribution of manganese (black) and iron (yellow-red) oxides around the circumference of the shaft.

Instead, we have abundant evidence for Mn formation which occurs at the interface between soil matrix and free air. Our published analysis of the mineral staining on the Dinaledi Chamber skeletal material demonstrates not only the presence of black spots as noted by Thackeray[1], but also tide marks of both Mn and Fe minerals on many of the specimens[2]. An example of these tide marks is shown in Figure 3. Lichens do not create such tide marks. These patterns form at the interface of soil and free air, and are a reflection of the function of the relative position of individual bones (or conjoin fragments) in the soil profile.[2] Such patterns of mineral deposition are most parsimoniously explained by bone to soil matrix contact[3,14] and are inconsistent with lichen growth.

So what evidence points Thackeray[1] toward lichens? Thackeray's citation for lichen involvement is his own technical note[15] in which he suggested a possible link between manganese dioxide staining and lichen growth on hominin crania from Sterkfontein and Swartkrans. This note, similar to his current contribution, posits merely that 'the spotty, discontinuous distribution of young lichen thalli appears to be analogous to the spotty distribution of MnO_2 on at least some hominid crania'[3(p.28)]. To support this hypothesis, it is incumbent upon Thackeray to quantify this resemblance in some way and show that it is unlikely to occur by chance. We have no opinion about whether lichen has deposited Mn on surface rocks such as those illustrated from Kromdraai (Figures 2 and 3 in Thackeray[1]). However, the lichen thalli illustrated in Thackeray's photographs do not have similar sizes to the Mn spots he has chosen to illustrate together with them. Nor do they have a similar density distribution. A very simple mode of quantification is to count the number of lichen thalli per square centimetre and the number of Mn spots; in the illustrated examples, these appear to be completely dissimilar. Further, Thackeray might have attempted to quantify the overall shape of lichen colonies and the overall shape of Mn staining. From his illustrations, we do not detect any such similarity except in the most general sense, in that they both have some dendritic portions, and thus any association with lichen may be a case of visual pareidolia.[16] As such, any possible causal link awaits empirical testing and validation. Thackeray admits this is the case, indicating that '…long-term experimental work on lichen on bone substrates in the Cradle of Humankind is planned for a 10-year period, in and around cave environments'[1(p.5)].

In addition to lichen, Thackeray suggests[1(p.5)] that snails and beetles could indicate the presence of leaf litter (on which they feed), which would be found near the cave entrance and therefore in a situation with diffuse light. However, as noted by Dirks and colleagues[13], snails have been recognised to colonise dark caves to considerable depth[17], and are far from restricted in diet to just plant matter[18]. *Gulella* sp. and *Euonyma varia* (Connolly, 1910) occur in large numbers in the Rising Star Cave system, including in deep chambers in the dark zone, and may well be responsible for some of the bone surface modifications on the fossils. Subulinid snails (including *Euonyma*) are largely omnivorous and are thus not dependent on green plant material as a food source (Herbert D 2016, personal communication, January 28), while *Gullela* spp. are carnivorous and feed mostly on other snails.[18]

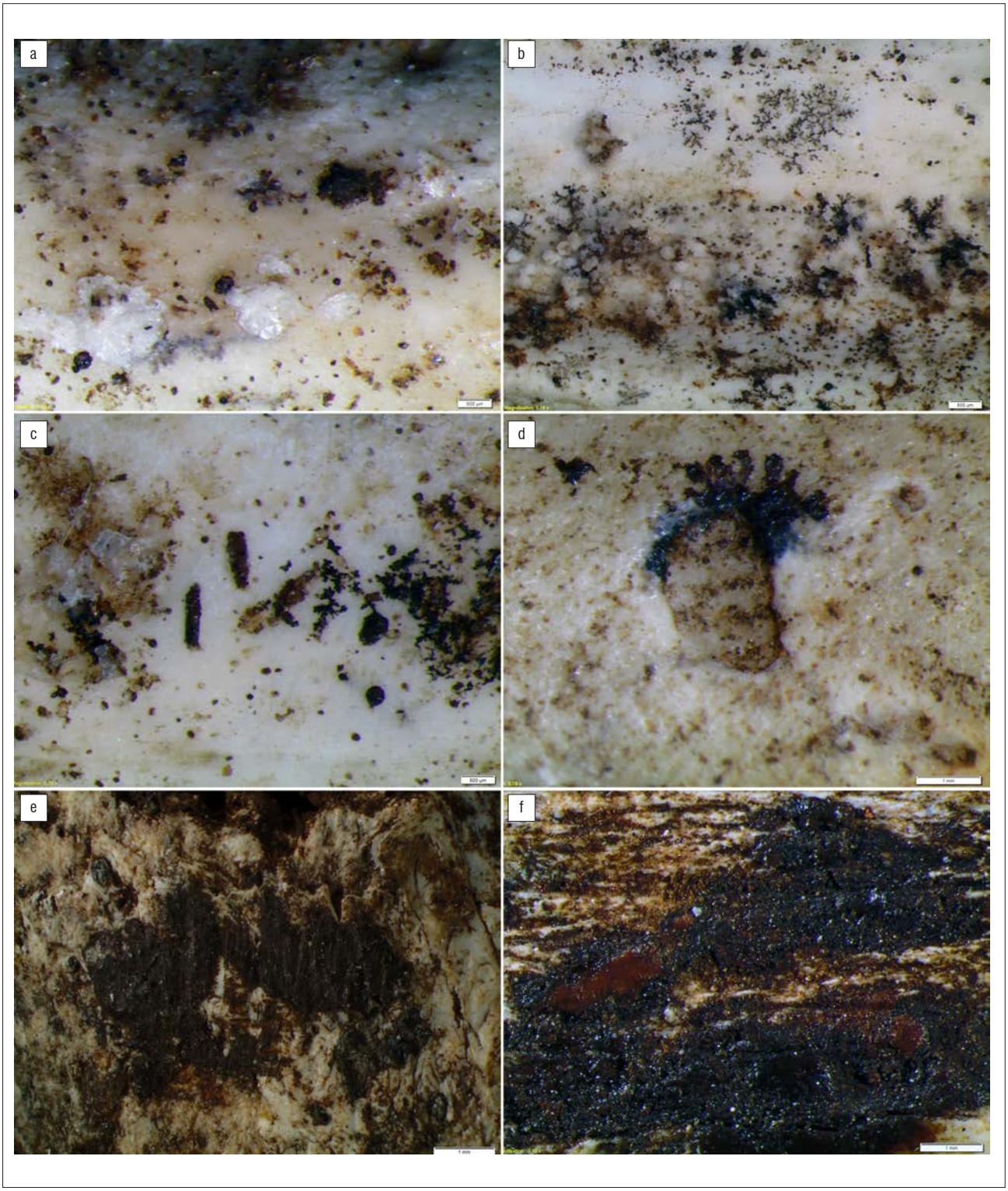
Figure 2: Patterns of mineral staining affecting tibia U.W. 101-1070. Note differential mineral staining patterns between conjoined fragments at the distal end.

Figure 3: Specimen U.W. 101–419 Cranium A(1) displaying tide lines of mineral staining which extends across different vault fragments. Tide lines mark a contact boundary between the bone surface and surrounding sediment, and indicate the resting orientation of the bone during precipitation of the stains.

Thus the notion that snails are light-dwelling, leaf-litter feeders is incorrect, and bone surface modification by snails cannot be used to imply a close entrance to the surface or redeposition of the hominin fossil material from a location near light. Furthermore, much of the radula damage observed on the bones of *H. naledi* occurs *after* they were mineralised – therefore most invertebrate damage was probably produced inside the dark Dinaledi Chamber on bones already covered in coatings of manganese and iron oxide deposits (Figure 4).

Our work on the taphonomy and geology of the Dinaledi Chamber has been multidisciplinary, and we re-emphasise (as in Dirks et al.[2]) that any hypothesis must be consistent with all lines of evidence including geological, geochemical and taphonomic data. Thackeray proposes that an alternative opening to the Dinaledi Chamber, capable of transmitting diffuse light from the surface to this location more than 30 m underground, may have existed at the time the fossils were deposited. He cites the commentary of Val[19], who likewise suggested that the fossils were once close to the surface in an environment with some light. Neither critic considers the strong geological or sedimentological evidence against such a scenario.[13] As detailed in Dirks and colleagues[2], the basic stratigraphic development of the Dinaledi Chamber comprises two facies subdivided into three stratigraphic units. Of these, hominin remains are found in Unit 2 remnants, with the bulk of the assemblage derived from Unit 3 – which accumulated along the floor of the chamber and is composed of largely unconsolidated sediment derived from weathering and erosion of Units 1 and 2.[2] Sedimentological analyses indicate that the clay-rich sediments making up these units were derived from in-situ weathering, and from exogenous clays and silts, which entered the chamber through fractures that prevented passage of coarser-grained materials. Thus the infill of the Dinaledi Chamber is the end product of a series of filters or traps, which winnowed out all large-grained sediments or clastic material, en route to final deposition within the chamber. The sediment inside these fossil-bearing units of the Dinaledi Chamber is significantly different in particle size and composition[2] from the neighbouring Dragon's Back Chamber, which is presently the only route from the Dinaledi Chamber toward the surface. If the fossils had been deposited at a time when there existed a larger opening into the Dinaledi Chamber, a secondary opening, or any substantially more open route to the surface, the sediment would not have these properties. Further, if the entrance to the Dinaledi Chamber had been illuminated and, therefore, accessible to surface fauna, or if there had not been special selection for only hominins, the bones of non-hominin fauna should be evident within the assemblage, in addition to other taphonomic markers of sub-aerial exposure, which are absent.[2,13]

In summary, we find that Thackeray's hypothesis of lichen deposition of Mn upon *H. naledi* fossil material is inconsistent with available evidence. Published evidence indicates that access to the Dinaledi Chamber was restricted to a single species of large-bodied animal, deposited over time, to the exclusion of all other animal forms during the period of the *H. naledi* depositional event.[2] Any alternative model must allow fleshed and articulated remains to enter the chamber, including articulated hands and feet[20,21], which are areas of anatomy that in articulation are unlikely to survive even short transport unless held together by soft tissues at the time of transport[22,23]. Furthermore, any other model must be restrictive enough that very limited externally derived sediments or organic material entered the chamber. Given these facts, we see nothing presented by Thackeray[1], or other commentators[19], that disproves the deliberate body disposal hypothesis we have put forward[2].



Scales in (a) and (b) = 500 µm; (c)–(f) = 1 mm.

Figure 4: (a) Black manganese deposits above and below petals of calcium carbonate on the bone surface (U.W. 101-40b). Dendrites penetrate bone and tooth surfaces, sometimes staining them grey. (b) Black manganese and orange iron oxide deposits on specimen U.W. 101-35. Note the dendritic pattern and tiny balls of calcium carbonate and manganese on the lower half of the bone. (c) Iron oxide deposit associated with balls of manganese oxide and modern frass in the centre of specimen U.W. 101-40c. The occurrence of iron and manganese dendrites suggest a microbial origin for the pattern of deposition. (d) Manganese deposit presenting as an amorphous ink-like stain on U.W. 101-965, which appears to have been partially removed by a gastropod. (e) Thick manganese coating on top of an iron oxide deposit on specimen U.W. 101-711. (f) Second-generation iron oxide deposit on a manganese coating that overlies an iron oxide one on specimen U.W. 101-312.

References

1. Thackeray FJ. The possibility of lichen growth on bones of *Homo naledi*: Were they exposed to light? S Afr J Sci. 2016;112(7/8), Art. #a0167, 5 pages. http://dx.doi.org/10.17159/sajs.2016/a0167
2. Dirks PHGM, Berger LR, Roberts EM, Kramers JD, Hawks J, Randolph-Quinney PS, et al. Geological and taphonomic context for the new hominin species *Homo naledi* from the Dinaledi Chamber, South Africa. eLife. 2015;4, e09561, 37 pages. http://dx.doi.org/10.7554/eLife.09561
3. Fernandez-Jalvo Y, Andrews P. Discolouration and staining. In: Fernandez-Jalvo Y, Andrews P, editors. Atlas of taphonomic identifications. London: Springer; 2016. p. 155–68. http://dx.doi.org/10.1007/978-94-017-7432-1_5
4. Millard AR, Hedges REM. A diffusion-adsorption model of uranium uptake by archaeological bone. Geochimic Cosmochim Ac. 1999;60:2139–2152. http://dx.doi.org/10.1016/0016-7037(96)00050-6
5. Pate FD, Hutton JT, Norrish K. Ionic exchange between soil solution and bone: Toward a predictive model. Appl Geochem. 1989;4:303–316. http://dx.doi.org/10.1016/0883-2927(89)90034-6
6. Karkanas P, Bar-Yosef O, Goldberg P, Weiner S. Diagenesis in prehistoric caves: The use of minerals that form in situ to assess the completeness of the archaeological record. J Archaeol Sci. 2000;27(10):915–929. http://dx.doi.org/10.1006/jasc.1999.0506
7. Nhauro G. Characterization of the elemental deposits in fossils from the Cradle of Humankind in South Africa and modern bones from the same geological area. Johannesburg: University of the Witwatersrand; 2010.
8. López-González F, Grandal-d'Anglade A, Vidal-Romaní JR. Deciphering bone depositional sequences in caves through the study of manganese coatings. J Archaeol Sci. 2006;33(5):707–717. http://dx.doi.org/10.1016/j.jas.2005.10.006
9. Diem D, Stumm W. Is dissolved Mn(II) being oxidized by O_2 in absence of Mn-bacteria or surface catalysts? Geochim Cosmochim Ac. 1984;48:1571–1573. http://dx.doi.org/10.1016/0016-7037(84)90413-7
10. Tebo BM, Johnson HA, McCarthy JK, Templeton AS. Geomicrobiology of manganese(II) oxidation. Trends Microbiol. 2005;13:421–428. http://dx.doi.org/10.1016/j.tim.2005.07.009
11. Post JE. Manganese oxide minerals: Crystal structures and economic and environmental significance. Proc Natl Acad Sci USA. 1999;96:3447–3454. http://dx.doi.org/10.1073/pnas.96.7.3447
12. Pfretzschner HU. Fossilization of Haversian bone in aquatic environments. CR Palevol. 2004;3:605–616. http://dx.doi.org/10.1016/j.crpv.2004.07.006
13. Dirks PHGM, Berger LR, Hawks J, Randolph-Quinney PS, Backwell LR, Roberts EM. Comment on 'Deliberate body disposal by hominins in the Dinaledi Chamber, Cradle of Humankind, South Africa?' [J. Hum. Evol. 96(2016)145-148]. J Hum Evol. 2016;96:149–153. http://dx.doi.org/10.1016/j.jhevol.2016.04.007
14. Pokines JT, Baker JE. Effects of burial environment on osseous remains. In: Pokines J, Symes S, editors. Manual of forensic taphonomy. Boca Raton, FL: CRC Press; 2013. p. 73–114. http://dx.doi.org/10.1201/b15424-6
15. Thackeray JF, Sénégas F, Laden G. Manganese dioxide staining on hominid crania from Sterkfontein and Swartkrans: Possible associations with lichen. S Afr J Sci. 2005;101:28.
16. Liu J, Li J, Feng L, Li L, Tian J, Lee K. Seeing Jesus in toast: Neural and behavioural correlates of face pareidolia. Cortex. 2014;53:60–77. http://dx.doi.org/10.1016/j.cortex.2014.01.013
17. Weigand A. New *Zospeum* species (Gastropoda, Ellobioidea, Carychiidae) from 980 m depth in the Lukina Jama-Trojama cave system (Velebit Mts., Croatia). Subterranean Biol. 2013;11:45–53. http://dx.doi.org/10.3897/subtbiol.11.5966
18. Herbert D, Kilburn D. Field guide to the land snails and slugs of eastern South Africa. Durban: Natal Museum; 2004.
19. Val A. Deliberate body disposal by hominins in the Dinaledi Chamber, Cradle of Humankind, South Africa? J Hum Evol. 2016;96:145–148. http://dx.doi.org/10.1016/j.jhevol.2016.02.004
20. Kivell TL, Deane AS, Tocheri MW, Orr CM, Schmid P, Hawks J, et al. The hand of *Homo naledi*. Nature Commun. 2015;6, Art. #8431, 9 pages. http://dx.doi.org/10.1038/ncomms9431
21. Kruger A, Randolph-Quinney PS, Elliott M. Multimodal spatial mapping and visualisation of the Dinaledi Chamber and Rising Star Cave. S Afr J Sci. 2016;112(5/6), Art. #2016-0032, 11 pages. http://dx.doi.org/10.17159/sajs.2016/20160032
22. Pokines JT. Faunal dispersal, reconcentration, and gnawing damage to bone in terrestrial environments. In: Pokines J, Symes S, editors. Manual of forensic taphonomy. Boca Raton, FL: CRC Press; 2013. p. 201–248. http://dx.doi.org/10.1201/b15424-10
23. Galloway A. The process of decomposition: A model from the Arizona-Sonoran Desert. In: Haglund WD, Sorg MH, editors. Forensic taphonomy: The postmortem fate of human remains. London: CRC Press; 1997. p. 139–150.

AUTHORS:
Isabelle J. Ansorge[1]
Geoff Brundrit[1]
Jean Brundrit[2]
Rosemary Dorrington[3]
Sarah Fawcett[1]
David Gammon[4]
Tahlia Henry[1]
Juliet Hermes[5]
Beate Hölscher[5]
Jethan d'Hotman[5]
Ian Meiklejohn[6]
Tammy Morris[1,5]
Izidine Pinto[7]
Marcel du Plessis[1]
Raymond Roman[1]
Clinton Saunders[8]
Fannie W. Shabangu[9]
Marc de Vos[10]
David R. Walker[11]
Gavin Louw[12]

AFFILIATIONS:
[1]Department of Oceanography, University of Cape Town, Cape Town, South Africa
[2]Department of Fine Art, University of Cape Town, Cape Town, South Africa
[3]Department of Biochemistry and Microbiology, Rhodes University, Grahamstown, South Africa
[4]Department of Chemistry, University of Cape Town, Cape Town, South Africa
[5]South African Environmental Observation Network, Phalaborwa, South Africa
[6]Department of Geography, Rhodes University, Grahamstown, South Africa
[7]Climate Systems Analysis Group, University of Cape Town, Cape Town, South Africa
[8]Sound and Vibration Research Group, Stellenbosch University, Stellenbosch, South Africa
[9]Fisheries Management, Department of Agriculture, Forestry and Fisheries, Roggebaai, South Africa
[10]Marine Research Section, South African Weather Service, Cape Town, South Africa
[11]Department of Conservation and Marine Studies, Cape Peninsula University of Technology, Cape Town, South Africa
[12]Oceans and Coasts Division, Department of Environmental Affairs, Roggebaai, South Africa

CORRESPONDENCE TO:
Isabelle Ansorge

EMAIL:
Isabelle.Ansorge@uct.ac.za

KEYWORDS:
marine science; National Development Plan; SA *Agulhas II*; postgraduates; teaching and research platform

HOW TO CITE:
Ansorge IJ, Brundrit G, Brundrit J, Dorrington R, Fawcett S, Gammon D, et al. SEAmester – South Africa's *first* class afloat. S Afr J Sci. 2016;112(9/10), Art. #a0171, 4 pages. http://dx.doi.org/10.17159/sajs.2016/a0171

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

SEAmester – South Africa's *first* class afloat

The Department of Science and Technology's (DST's) 10-year Global Change Grand Challenge programme requires platforms to 'attract young researchers to the region and retain them by exciting their interest in aspects of global change; while developing their capacity and professional skills in the relevant fields of investigation'[1]. In addition, in July 2014, President Zuma officially launched Operation Phakisa and announced that a key target of this Oceans Economy initiative would be 'for the Department of Higher Education and Training to drive alignment between theoretical and workplace learning'[2].

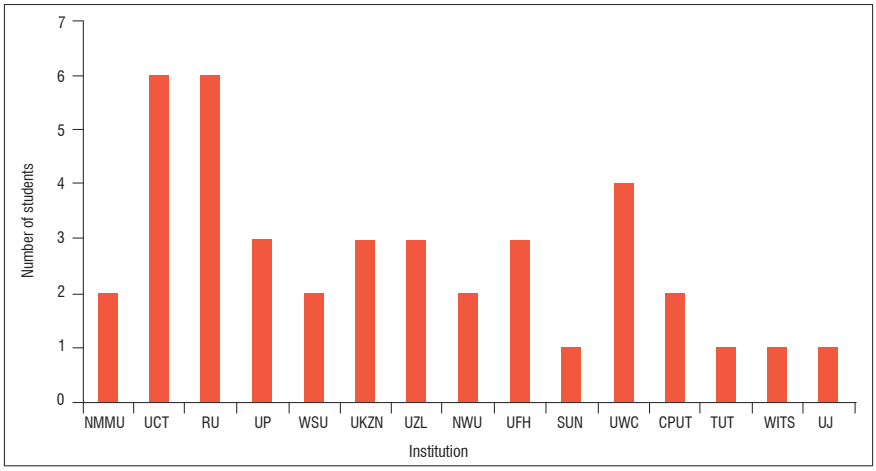
SEAmester – South Africa's recently established Class Afloat – achieves just that. SEAmester introduces marine science as an applied and cross-disciplinary field to students who have shown an affinity for core science disciplines. It identifies with government's National Development Plan[3] on education, training and innovation – critical to South Africa's long-term development and investment in this sector.

SEAmester has a long-term vision aimed at building capacity within the marine sciences by coordinating and fostering cross-disciplinary research projects and achieving this goal through a highly innovative programme. The strength of SEAmester is that postgraduate students combine theoretical classroom learning with the application of this knowledge through ship-based, and more importantly, hands-on research. The state-of-the-art research vessel, SA *Agulhas II*, provides an ideal teaching and research platform for this programme; its size, comfort and shipboard facilities allow large groups of students and lecturers to productively interact over a period of 10 days.

Introduction

Marine science is a highly competitive environment. The need to improve the cohort of South African postgraduates, who would be recognised both nationally and internationally for their scientific excellence, is crucial. It is possible to attract students early on in their careers to this discipline via cutting-edge science, technology and unique field experiences. Through the engagement of students with real-life experiences such as SEAmester, universities supporting marine science postgraduate degree programmes can attract a sustainable throughput of numerically proficient students. By achieving a more quantitative and experienced input into our postgraduate degree programmes, we will, as a scientific community, greatly improve our long-term capabilities to accurately measure, model and predict the impacts of current climate change scenarios.

The short-term goal is to attract and establish a cohort of proficient marine and atmospheric science graduates who will contribute to filling the capacity needs of South African marine science as a whole. The SEAmester programme, by involving researchers from across all the relevant disciplines and tertiary institutions, provides an opportunity to build a network of collaborative teaching within the marine field. In doing so, these researchers will foster and strengthen new and current collaborations between historically white and black universities (Figure 1). The long-term objective of SEAmester is to build critical mass within the marine sciences to ensure sustained growth of human capacity in marine science in South Africa – aligning closely with the current DST Research and Development strategies and the Operation Phakisa Oceans Economy initiative.



NMMU, Nelson Mandela Metropolitan University; UCT, University of Cape Town; RU, Rhodes University; UP, University of Pretoria; WSU, Walter Sisulu University; UKZN, University of KwaZulu-Natal; UZL, University of Zululand; NWU, North West University; UFH, University of Fort Hare; SUN, Stellenbosch University; UWC, University of the Western Cape; CPUT, Cape Peninsula University of Technology; TUT, Tshwane University of Technology; WITS, University of the Witwatersrand; UJ, University of Johannesburg.

Figure 1: A histogram showing the range of universities represented on board the 2016 SEAmester cruise.

A nationwide call to postgraduate students

On 1 June 2016, a call through the South African Network for Coastal and Oceanic Research was opened for postgraduate students studying marine sciences or a similar discipline to join SEAmester. At the closure of the call on 8 June, SEAmester had attracted 132 applications from 18 universities across South Africa and Namibia. Each application was evaluated on academic merit. A total of 40 applications from students at honours, MSc and PhD levels was selected. A core element in the selection process was to ensure that SEAmester was transformative and inclusive, and hence opportunities were offered to students from institutions that traditionally do not offer ocean-going research in their curriculum (Figure 1).

In addition, 15 researchers and academics working in the Departments of Chemistry, Fine Art, Environmental and Geographical Science and Oceanography at UCT, Departments of Microbiology and Geography at Rhodes University, Department of Engineering at Stellenbosch University, Biodiversity Division at the Department of Agriculture, Forestry and Fisheries, Department of Marine Science at the Cape Peninsula University of Technology, Birdlife SA, South African Weather Service and South African Environmental Observation Network (SAEON) Egagasini Node joined the SA *Agulhas II* cruise as part of the lecturing cohort.

SEAmester provides students with a choice of academic stream

On the recent July 2016 cruise, the module 'Oceans in a Changing Climate' appealed to students wishing to learn more about the impact that the interactions between physical, chemical and biological ocean processes and change are having on marine ecosystems. The more technically and data-minded students chose to learn techniques in data collection, marine instrumentation and the technology behind the ship-based physical and biological instrumentation under 'Tools of the Trade'. Underpinning both streams was a course in 'Ocean Dynamics', given to all students in order to provide a better understanding of the global ocean and atmospheric processes, as well as how the oceans contribute to earth's climate by storing and transporting heat and salt between ocean basins (Figure 2).

Learning during the SEAmester cruise included intensive daily classroom lectures and assignments (Figure 2). In addition to classes and course-work, students provided research support to scientists working on the Agulhas System Climate Array (ASCA) programme and obtained real hands-on training (Figure 3). ASCA is a partnership between South African and international marine science institutes and is coordinated by SAEON. The focus of the array is long-term monitoring of heat, salt and volume transport in the Agulhas Current. Given the importance of the Agulhas Current in global and regional climate, the long-term nature of the ASCA programme, the variety of instruments used and the potential for further studies with the data, ASCA provided an ideal learning platform for the inaugural SEAmester cruise.

Deck-based training consisted of working with oceanographic, atmospheric and biological ship-based instrumentation, as well as underway measurements and autonomous devices such as Argo floats. Training on ocean data analysis was provided, and technical experience was gained in data management such as salinity and oxygen sensor calibrations. In addition, student groups worked on a final project presentation related to their area of study interest and the scientific research undertaken during the voyage.

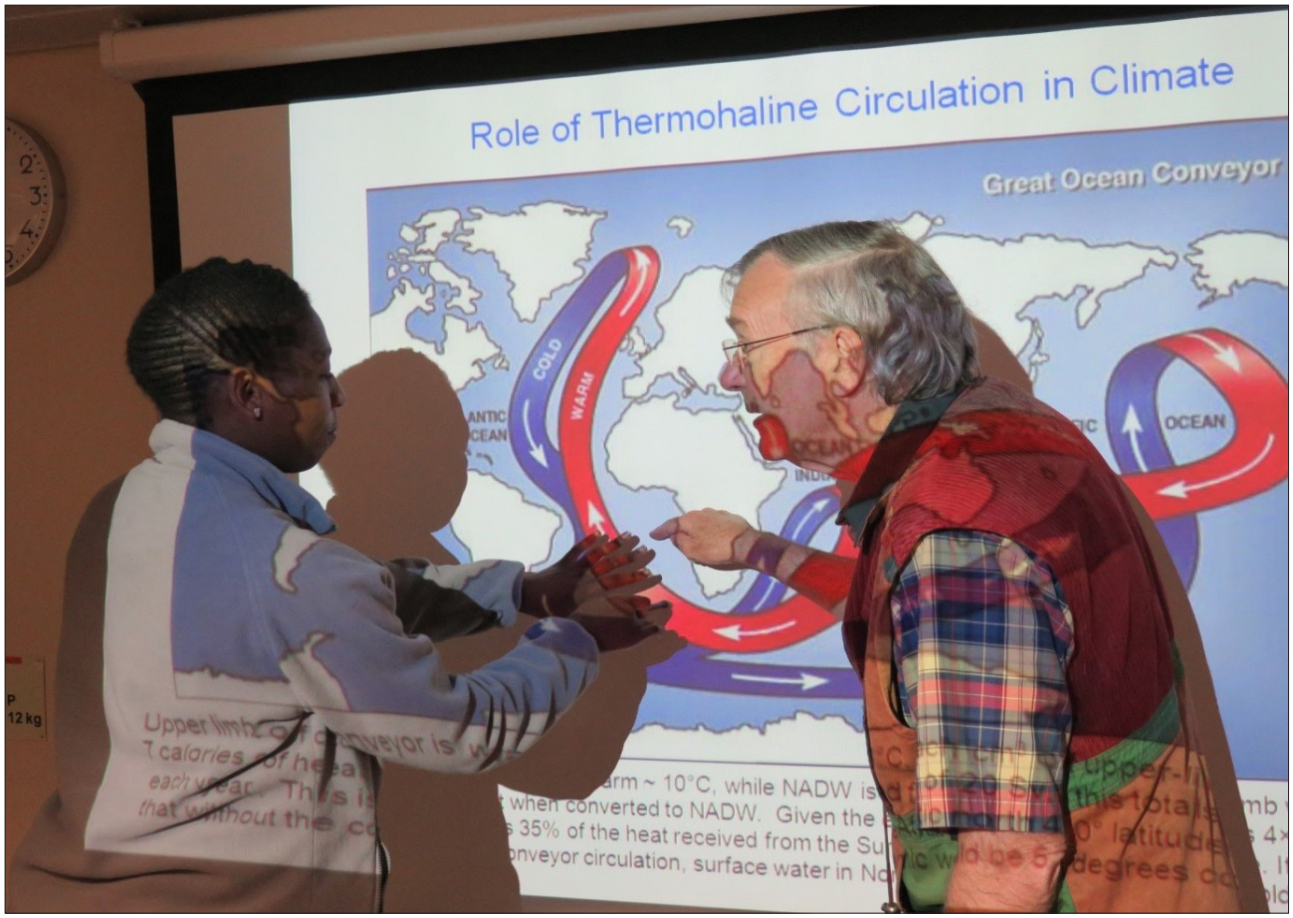


Photo: Jean Brundrit

Figure 2: SEAmester student Miss Tebogo Masebe from Rhodes University listening to Emeritus Professor Geoff Brundrit explain the role of the Global Conveyor Belt. Emeritus Professor Brundrit is a former Head of Oceanography at UCT and Special Advisor on ocean climate change to the Department of Environmental Affairs.

Photo: Jean Brundrit

Figure 3: SEAmester students getting the CTD (conductivity, temperature and depth) device ready for water sample collection at the next station. From left to right: Stephen Peel (UCT), Hermann Luyt (SUN), Sizakele Sibanda (UCT), Belinda Nhesvure (UFH), Thoko More (TUT) and Ntombi Nxiba (CPUT).

How important was SEAmester to each student?

SEAmester is a unique shipboard programme that integrates interdisciplinary coursework, hands-on ship-based experiences (Figure 3) and interaction with leading South African marine researchers. By aligning with the scientific ASCA programme, SEAmester also allowed students to collect data in an oceanic region of global importance and to be part of an international programme with data standards and protocols.

The long-term goal is to establish SEAmester as an annual premier tertiary programme for multidisciplinary marine research, teaching and training in southern Africa. Its continued role will be to attract numerically proficient students into the marine sciences by coordinating and fostering cross-disciplinary and interfaculty research projects and curricula, while training students at sea. Areas of particular development and capacity building within this programme will include marine meteorology; physical, biological and operational oceanography; and an ecosystem approach to marine biodiversity and biogeochemical studies. The success of the recent 2016 SEAmester cruise is evident in the following postgraduate student comments from the SEAmester evaluation form:

- *This programme has been the best experience in my life. It changed my outlook on how to gather knowledge, where to do research, how to think logically as well as to be creative and intuitive.*
- *SEAmester has made me realise that I have a passion for ocean-based research.*
- *I loved the way the content of SEAmester has helped me to link up all the theory that I have learnt over the years.*
- *I have been changed positively by this course – I am more motivated to tackle my studies.*
- *It was a time of my life that I will never ever forget.*
- *SEAmester has been a life changing experience for me.*

We believe that SEAmester offers an unparalleled opportunity to live and work in the marine environment, and in doing so leaves a tangible and lasting impression that postgraduate students are able to make a meaningful contribution to the field of marine science. The government's National Development Plan identifies education, training and innovation as being at the forefront of South Africa's long-term development, and specifically states[3]: 'Inadequate capacity will constrain knowledge production and innovation unless effectively addressed.' The central strategy behind SEAmester is not only of a training programme that aligns with core DST objectives, but more importantly, a programme designed to fill a burgeoning gap in the capabilities of our current cohort of marine students. It aims to build critical mass within the marine sciences, so as to ensure sustained growth in this field within South Africa and beyond.

Greater awareness of the ocean's physical, biogeochemical and ecological response to climate change, highlighted through ship-board experiences, has already started to inspire and attract students into

the marine sciences – a critical step if a new generation of southern African marine scientists with a far higher calibre in the sciences are to be trained, and a key objective of Operation Phakisa is to be realised. The success of the recent cruise on board the SA *Agulhas II* has confirmed to the scientific community that SEAmester – South Africa's Class Afloat – is able to achieve just that.

Acknowledgements

SEAmester is funded through the National Research Foundation (South Africa) and DST and logistically supported by the Department of Environmental Affairs (DEA). The Agulhas Systems Climate Array (ASCA) is funded through the DST/SAEON/DEA/NIOZ/RSMAS/NSF partnership. The successful completion of the survey would not have been possible without the invaluable assistance of the officers and crew of the SA *Agulhas II*. In particular, we thank Captain Gavin Syndercombe for his continued insight, support and guidance throughout the planning and execution of the SEAmester cruise. We also thank Dr Gilbert Siko and Tracy Klareenbeck from the South African National Antarctic Programme and Bigboy Joseph and Nish Devanunthan from Antarctica and Islands Logistics (DEA) for their continued support. Our thanks also go to the senior technical electronics officer Angelo Marco for his invaluable assistance throughout the entire voyage. And finally, we thank all the cruise participants involved in ASCA who helped to collect data and provided additional deck training to the students.

References

1. South African Department of Science and Technology (DST). The 10-year plan for science and technology (2008–2018). Pretoria: DST; 2008. Available from: http://www.esastap.org.za/download/sa_ten_year_innovation_plan.pdf
2. Address by His Excellency President Jacob Zuma at the Operation Phakisa: Unlocking the Economic Potential of the Ocean Economy, Open Day International Convention Centre; 2014 October 15; Durban, South Africa. Available from: http://www.thepresidency.gov.za
3. National Development Plan 2030: Our future – make it work. Pretoria: The Presidency; 2012. Available from: http://www.gov.za/sites/www.gov.za/files/NDP-2030-Our-future-make-it-work_r.pdf

South Africa in the Antarctic Circumnavigation Expedition: A multi-institutional and interdisciplinary scientific project

AUTHORS:
Issufo Halo[1]
Rosemary Dorrington[2]
Thomas Bornman[3,4]
Stephanie de Villiers[5,6]
Sarah Fawcett[7]

AFFILIATIONS:
[1]Department of Conservation and Marine Sciences, Cape Peninsula University of Technology, Cape Town, South Africa
[2]Department of Biochemistry and Microbiology, Rhodes University, Grahamstown, South Africa
[3]South African Environmental Observation Network, Elwandle Coastal Node, Port Elizabeth, South Africa
[4]Department of Botany, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa
[5]Oceans and Coasts Research, Department of Environmental Affairs, Cape Town, South Africa
[6]Department of Chemistry, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa
[7]Department of Oceanography, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Sarah Fawcett

EMAIL:
sarah.fawcett@uct.ac.za

KEYWORDS:
Subantarctic ecosystems; Antarctic Circumpolar Current; biological pump; microbial diversity; macronutrient cycling

HOW TO CITE:
Halo I, Dorrington R, Bornman T, De Villiers S, Fawcett S. South Africa in the Antarctic Circumnavigation Expedition: A multi-institutional and interdisciplinary scientific project. S Afr J Sci. 2016;112(9/10), Art. #a0173, 4 pages. http://dx.doi.org/10.17159/sajs.2016/a0173

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

The polar regions are more critically affected by climate change than any other region on our planet.[1,2] On the Antarctic continent and in its surrounding oceans, the effects of climate change are likely to be dramatic,[3] and include large-scale catastrophic ice melt, loss of habitat and biodiversity, and global sea level rise. The 'Southern Ocean' refers to the region where Atlantic, Indian and Pacific Ocean waters come together to encircle Antarctica. These waters connect the different ocean basins by linking the shallow and deep limbs of the global ocean current system ('overturning circulation') and play a critical role in storing and distributing heat and carbon dioxide (CO_2). The Southern Ocean thus regulates not only the climate of the Antarctic, but of the entire earth system.[1,4] By extension, the capacity of the global ocean to ameliorate earth's changing climate is strongly controlled by the Southern Ocean.

Marine phytoplankton (microscopic plants inhabiting the sunlit upper ocean) convert CO_2 (an inorganic form of carbon) dissolved in surface waters into organic carbon through photosynthesis. This organic carbon fuels upper trophic levels such as fish, mammals and birds, and a portion sinks into the deep ocean where it remains stored for hundreds to thousands of years. This mechanism, which lowers the atmospheric concentration of CO_2, is termed the 'biological pump'.[5] The efficiency of the global ocean's biological pump is currently limited by the Southern Ocean, where the macronutrients (nitrate and phosphate) required for photosynthesis are never fully consumed in surface waters. In theory, increased consumption of these nutrients could drive higher organic carbon removal to the deep ocean, enhancing the oceanic uptake of atmospheric CO_2. Indeed, more complete consumption of Southern Ocean nutrients is a leading hypothesis for the decrease in atmospheric CO_2 that characterised the ice ages.[6]

Despite the global importance of the Southern Ocean, knowledge of the controls on and interactions among the physical, chemical and biological processes operating in Antarctic ecosystems is limited, largely because of a scarcity of in-situ observational data, compounded by the challenge of integrating siloed scientific fields. Given predictions that diverse aspects of Southern Ocean physics and carbon biogeochemistry are likely to change in the coming decades, a transdisciplinary approach to studying Antarctic systems is critical.

The ACE project

Driven by the urgent need to improve our understanding of the Southern Ocean and its ecosystems, an international, competitive open call was announced in 2015 for research proposals to participate in an oceanographic voyage around Antarctica (Figure 1). This project – the Antarctic Circumnavigation Expedition (ACE) – plans to follow the path of the largest oceanographic feature of the Southern Ocean, the eastward propagating, wind-driven Antarctic Circumpolar Current (ACC), which mixes oceanic properties across ocean basins and connects numerous Subantarctic island systems. The waters surrounding these islands experience high primary productivity fuelled by nutrients supplied through island run-off and seal and bird activity.[7] The islands also play an important role in the creation and behaviour of eddies[8], which can trap oceanic properties and transport them long distances[9,10], enhancing the role of the ACC as a 'conveyor belt' that connects ocean basins and biological systems (Figure 2).

The goal of ACE, stated in the ACE Project Call for Proposals, is 'to offer international teams of distinguished scientists an outstanding and unique opportunity to study the marine and terrestrial environment of the Subantarctic ecosystem', in a single summer cruise that will start and end in Cape Town (20 December 2016 to 18 March 2017). Over 100 project proposals were submitted to the call, from which 22 were selected by an international panel of experts convened to evaluate them according to criteria based on scientific excellence. Committee members included representatives from polar research institutes in Australia, France, Norway, the Russian Federation, South Africa, Switzerland and the United Kingdom. The 22 selected ACE research projects involve 55 marine and terrestrial researchers from 30 countries working in atmospheric science, biogeochemistry, climatology, glaciology, marine and terrestrial ecology, ocean engineering, oceanography and palaeoclimatology.

The ACE expedition is supported by Ferring Pharmaceuticals and is the maiden project of the newly created Swiss Polar Institute located at the École Polytechnique Fédérale de Lausanne (EPFL), the Swiss Institute of Forest, Snow and Landscape Research WSL, the Swiss Federal Institute of Technology in Zurich (ETHZ), the University of Bern, and Editions Paulsen. ACE will be conducted on board the Russian polar research vessel, *Akademik Treshnikov*, and will consist of three legs: Leg 1 is from Cape Town to Hobart, Leg 2 is from Hobart to Punta Arenas and Leg 3 is from Punta Arenas to Cape Town (Figure 1). During each cruise leg, the ship will stop at numerous Subantarctic islands located in the path of the ACC, allowing for terrestrial and coastal research in addition to the open ocean investigations that will be the focus of the ship's transects.

ACE major objectives and collaborators

Beyond the specific aims of the 22 scientific projects, the objectives of ACE, outlined in the Call for Proposals, are twofold: firstly, 'to enhance international relations and collaborations between countries', and secondly, 'to promote the interest of a new generation of young explorers in polar research'. To achieve these objectives, the implementation of the ACE project is being facilitated by the following partners in addition to the Swiss Polar Institute and EPFL: The Australian Antarctic Division, Institut Paul Émile Victor (France), the Norwegian Polar Institute, the Arctic and Antarctic Research Institute (the Russia Federation), the British Antarctic Survey (BAS), and South Africa's University of Cape Town (UCT), National Antarctic Programme (SANAP), National Research Foundation (NRF) and Department of Science and Technology (DST).

Source: ©EPFL

Figure 1: Proposed route for the Antarctic Circumnavigation Expedition (ACE), on board the R/V *Akademik Treshnikov*. The three legs of ACE are Cape Town to Hobart (Leg 1), Hobart to Punta Arenas (Leg 2) and Punta Arenas to Cape Town (Leg 3). The expected timeline for the cruise is 20 December 2016 to 18 March 2017.

South African contribution: ACE Project XII

South Africa is the only African country to have submitted a successful application to ACE, ACE Project XII: 'A multi-disciplinary, multi-resolution approach to understanding nutrient cycling and microbial diversity in changing Subantarctic ecosystems'. The principal investigator (PI) of this project is Dr Sarah Fawcett from the Department of Oceanography, UCT, with Prof. Rosemary Dorrington from the Department of Biochemistry and Microbiology, Rhodes University, as co-PI. The project involves an additional three South African co-PIs and their collaborators: Dr Thomas Bornman, Elwandle Coastal Node, South African Environmental Observation Network (SAEON) and Nelson Mandela Metropolitan University (NMMU); Dr Stephanie de Villiers, Oceans and Coasts Research, Department of Environmental Affairs (DEA) and NMMU; and Dr Issufo Halo, Department of Conservation and Marine Sciences, Cape Peninsula University of Technology (CPUT). International collaborators from the Department of Geosciences, Princeton University, USA, are also involved.

The broad goal of ACE Project XII is to use microbial diversity (where 'microbial' refers to phytoplankton, bacteria and zooplankton) and metabolic activity in conjunction with measured chemical and physical parameters to develop an integrated model of the Subantarctic island systems in order to better understand their role in Southern Ocean productivity. Investigations of microbe–nutrient interactions will focus on the nitrogen (N) cycle because N exists in marine systems in numerous forms, is linked in a roughly constant ratio to carbon, and is primarily transformed by biology.[11] N is supplied to phytoplankton in different ways. For example, the ocean interior is filled with nitrate (NO_3^-), which is delivered to the sunlit surface ocean by the upward mixing of deep waters. Ammonium (NH_4^+) is produced in surface waters as a by-product of bacterial and zooplankton metabolic activity, and urea (an organic N form) is generated by birds and seals inhabiting the Subantarctic islands and washed into the open ocean by rain or snow-melt. Annually, the amount of phytoplankton growth fuelled by N sources supplied from outside the surface ocean ('new N', the phytoplankton growth upon which is termed 'new production') is balanced by a flux of organic N (and carbon) to the deep ocean ('export production').[12,13] The source and form of N consumed by phytoplankton is thus directly related to the capacity of an ecosystem to absorb CO_2.

To identify the N sources fuelling the island systems and open ACC, the N and oxygen (O) isotope ratios of nitrate and N isotopes of phytoplankton biomass will be measured and interpreted in the context of nitrate, ammonium and organic N concentration data. The N isotopes provide an indication of N source[14-17] and, when combined with measurements of nitrate O isotopes, enable the separation of co-occurring processes (such as nitrification by bacteria and nitrate assimilation by phytoplankton – two processes with very different implications for CO_2 cycling)[18-21]. Microbial community composition will be assessed through a combination of methods selected to maximise spatial coverage and resolution. These methods include: phytoplankton and zooplankton taxonomy; phytoplankton DNA/RNA microarrays that evaluate phytoplankton diversity and gene expression related to N- and carbon cycle processes[22-24]; phenotypic arrays that characterise the metabolic activity of bacterial communities[25,26]; Next Generation Sequencing for microbial diversity[27], and rRNA and cDNA for the identification of metabolically active microbial taxa. Finally, the physical, chemical, and biological data will be incorporated into ongoing high-resolution numerical modelling efforts to produce an integrated view of the Subantarctic systems.

In sum, ACE Project XII will use a multidisciplinary approach to address the following questions regarding Subantarctic microbes and their interactions with their physico-chemical environment: Who is there? What are they doing? Why are they doing it? What are the implications for Subantarctic nutrient cycling, ecosystem function, and CO_2 removal, today and in a warming world?

In addition to ACE Project XII, there is South African participation in seven other ACE projects (led by collaborator countries). South African co-PIs on these projects include: Prof. Ken Findlay, CPUT; Dr Sandy Thomalla, Council for Scientific and Industrial Research; Prof. Peter Ryan, UCT; Prof. Bettine Jansen van Vuuren and Dr Peter le Roux, University of Johannesburg; Prof. Marthaan Bester and Dr Nico de Bruyn, University of Pretoria; Dr Gwynneth Matcher, Rhodes University; and Dr Keith MacHutchon, Coastal Marine Technology Ltd.

Capacity development through ACE

The ACE project offers numerous opportunities for capacity development. Firstly, ACE is funding a Maritime University under the auspices of the Russian Geographical Society. This is an opportunity for postgraduate students from all over the world to participate in a ~25-day transit cruise from Bremerhaven (Germany) to Cape Town during which they will be introduced to the interdisciplinary field of applied marine science. Secondly, ACE Project XII will involve at least four MSc/PhD students and a postdoctoral researcher, to be based at any of the four affiliated South African universities; all co-PIs will contribute to student supervision, thus exposing students to scientific career possibilities within both government and academia. These students will participate in a 3-month multidisciplinary rotation programme between the Dorrington (biology), Fawcett (biogeochemistry), Halo (physics) and DEA/SAEON (marine ecosystem sampling for phytoplankton and zooplankton) groups, the goal of which is to equip them with a range of analytical skills and a broad view of marine systems. Thirdly, ACE Project XII brings together numerous sophisticated and novel analytical techniques, which, when coupled with research visits to international collaborator laboratories, will upskill the students involved and facilitate the transfer of scarce skills that can be applied far beyond this project. Fourthly, to increase the visibility of polar research in general and of the ACE expedition in particular, the ACE management team, together with local and national partners (UCT, DST, SANAP), are planning a series of public events in Cape Town to coincide with the December 2016 departure and March 2017 arrival of the *Akademik Treshnikov.* In addition, ACE Project XII plans to enhance public engagement in science and research through (1) educational efforts facilitated by SAEON's outreach network, which currently includes 13 schools and (2) activities that directly engage the public such as a cruise blog, public lectures, popular articles, media events and participation in national focus events such as National Science Week, National Marine Week and SciFest Africa.

Conclusion

It is anticipated that ACE Project XII (and South African participation in ACE in general) will have a nationwide impact. In addition to developing human capacity, ACE Project XII will yield a deeper fundamental understanding of nutrient cycling and microbial dynamics in the Antarctic, and provide information on ecosystem function in response to environmental drivers. The results will assist decision-makers in formulating appropriate environmental policies to lessen the vulnerability of Subantarctic systems; an example is the Prince Edward Islands, a South African Marine Protected Area (Figure 2).

The research proposed in ACE Project XII is fully aligned with the SANAP mission to 'increase understanding of the natural environment and life in the Antarctic [through] science'[28]. South Africa is also a member of the Commission for the Conservation of Antarctic Marine Living Resources, responsible for contributing to the global effort to conserve Southern Ocean living resources through monitoring and research. This contribution includes research aimed at better understanding the controls on primary productivity, and changes in productivity and carbon cycling related to emerging threats such as ocean warming and acidification. Such fundamental knowledge is key to managing the long-term impacts of global change on the feeding and population dynamics of marine species higher up the food chain, such as fish and top predators. Finally, given that Southern Ocean waters currently support globally significant fisheries, ACE is also relevant to the South African government's ongoing development initiative through Operation Phakisa to 'unlock the potential of South Africa's oceans'[29].

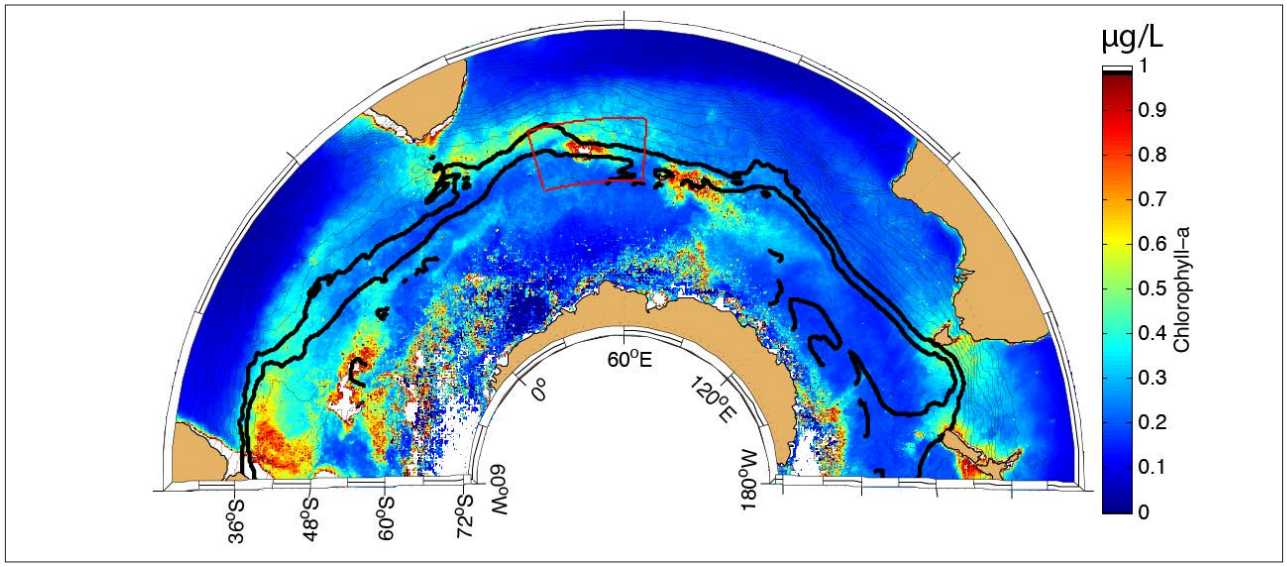


Figure 2: Chlorophyll-a concentrations (μg/L) in the Atlantic and Indian sectors of the Southern Ocean for the austral summer, derived from SeaWifs satellite data averaged for 1997–2004 (spatial grid resolution of 1 degree, nominal temporal resolution of 8 days). Chlorophyll-a is a green pigment used by all phytoplankton for light absorption during oxygenic photosynthesis; the distribution of surface ocean chlorophyll-a can thus be used as a proxy for phytoplankton productivity. The black contours represent streamlines of the mean dynamic topography, computed from the Rio-09 New CNES-CLS09 product and used as a proxy for large-scale geostrophic currents. Bold contours show the core of the Antarctic Circumpolar Current (ACC). The red box indicates the location of the remote South African Prince Edward Islands, where land–ocean interactions will be investigated as part of ACE Project XII. High microbial primary productivity can be inferred from the high chlorophyll-a concentrations around the islands and shallow banks and along the subtropical front.

Acknowledgements

We are grateful for the support of Isabelle Ansorge, South African ACE Technical Committee member, and Danièle Rod, Advisor to the Presidency at EPFL.

References

1. Anisimov OA, Vaughan DG, Callaghan TV, Furgal C, Merchant H, Prowse TD, et al. Polar regions (Arctic and Antarctic). In: Parry ML, Canzone OF, Palutikof JP, Van der Linden PJ, Hanson CE, editors. Climate change 2007: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2007. p. 653–685.
2. Cicerone R, Nurse P. Climate change: Evidence and causes. An overview from the Royal Society and the US National Academy of Sciences. Washington DC/ London: National Academy of Sciences/Royal Society; 2015. p. 1–36.
3. Lucas M, Gihwala K, Viskich M. How are Antarctica and the Southern Ocean responding to climate change? Quest. 2014;10(2):16–19.
4. Bigg GR, Jickells TD, Liss PS, Osborn TJ. The role of the oceans in climate. Int J Climatol. 2003;23:1127–1159. http://dx.doi.org/10.1002/joc.926
5. Volk T, Hoffert MI. Ocean carbon pumps: Analysis of relative strengths and efficiencies on ocean-atmospheric pCO_2 changes. In: Sundquist ET, Broecker WS, editors. The carbon cycle and atmospheric CO_2 natural variation. Vol 32. Washington DC: American Geophysical Union; 1985. p. 99–120. http://dx.doi.org/10.1029/GM032p0099
6. Sigman DM, Hain MP, Haug GH. The polar ocean and glacial cycles in atmospheric CO_2 concentration. Nature. 2010;466:47–55. http://dx.doi.org/10.1038/nature09149
7. Smith VR, Froneman PW. Nutrient dynamics in the vicinity of the Prince Edward Islands. In: Chown SL, Froneman PW, editors. The Prince Edward Islands: Land-sea interactions in a changing ecosystem. Stellenbosch: SUN PReSS; 2008. p. 165–179.
8. Durgadoo JV, Ansorge IJ, De Cuevas BA, Lutjeharms JRE, Coward AC. Decay of eddies at the South-West Indian Ridge. S Afr J Sci. 2011;107(11/12), Art. #673, 10 pages. http://dx.doi.org/10.4102/sajs.v107i11/12.673
9. Robinson AR, Lobel PS. Impacts of ocean eddies on coastal currents. In: Magaard L, Pujalet R, Gaynor V, editors. HOE: The Hawaiian Ocean Experiment. Proceedings of Aha Huliko'a, Hawaiian Winter Workshop; 1985 January 21–24; University of Hawaii, HI, USA. Honolulu, HI: Hawaii Institute of Geophysics; 1985. p. 325–334.
10. Chelton DB, Schlax MG, Samelson RM. Global observations of nonlinear mesoscale eddies. Prog Oceanogr. 2011;91:167–216. http://dx.doi.org/10.1016/j.pocean.2011.01.002
11. Zehr JP, Ward BB. Nitrogen cycling in the ocean: New perspectives on processes and paradigms. Appl Environ Microb. 2002;68:1015–1024. http://dx.doi.org/10.1128/AEM.68.3.1015-1024.2002
12. Dugdale RC, Goering JJ. Uptake of new and regenerated forms of nitrogen in primary production. Limnol Oceanogr. 1967;12:196–206. http://dx.doi.org/10.4319/lo.1967.12.2.0196
13. Eppley RW, Peterson BJ. Particulate organic matter flux and planktonic new production in the deep ocean. Nature. 1979;282:677–680. http://dx.doi.org/10.1038/282677a0
14. Altabet MA. Variations in nitrogen isotopic composition between sinking and suspended particles: Implications for nitrogen cycling and particle transformation in the open ocean. Deep-Sea Res. 1988;35:535–554. http://dx.doi.org/10.1016/0198-0149(88)90130-6
15. Fawcett SE, Lomas MW, Casey JR, Ward BB, Sigman DM. Assimilation of upwelled nitrate by small eukaryotes in the Sargasso Sea. Nat Geosci. 2011;4:717–722. http://dx.doi.org/10.1038/ngeo1265
16. Fawcett SE, Lomas MW, Ward BB, Sigman DM. The counterintuitive effect of summer-to-fall mixed layer deepening on eukaryotic new production in the Sargasso Sea. Global Biogeochem Cy. 2014;28:86–102. http://dx.doi.org/10.1002/2013GB004579
17. Treibergs LA, Fawcett SE, Lomas MW, Sigman DM. Nitrogen isotopic response of prokaryotic and eukaryotic phytoplankton to nitrate availability in Sargasso Sea surface waters. Limnol Oceanogr. 2014;59:972–985. http://dx.doi.org/10.4319/lo.2014.59.3.0972
18. Sigman DM, Granger J, DiFiore PJ, Lehmann MF, Ho R, Cane G, et al. Coupled nitrogen and oxygen isotope measurements of nitrate along the eastern North Pacific margin. Global Biogeochem Cy. 2005;19, Art. GB4022, 14 pages. http://dx.doi.org/10.1029/2005GB002458
19. Wankel SD, Kendall C, Pennington JT, Chavez FP, Paytan A. Nitrification in the euphotic zone as evidenced by nitrate dual isotopic composition: Observations from Monterey Bay, California. Global Biogeochem Cy. 2007;21, Art. GB2009, 13 pages. http://dx.doi.org/10.1029/2006GB002723
20. Fawcett SE, Ward BB, Lomas MW, Sigman DM. Vertical decoupling of nitrate assimilation and nitrification in the Sargasso Sea. Deep-Sea Res I. 2015;103:64–72. http://dx.doi.org/10.1016/j.dsr.2015.05.004
21. Smart SM, Fawcett SE, Thomalla SJ, Weigand MA, Reason CJC, Sigman DM. Isotopic evidence for nitrification in the Antarctic winter mixed layer. Global Biogeochem Cy. 2015;29:427–445. http://dx.doi.org/10.1002/2014GB005013
22. Ward BB, Eveillard D, Kirshtein JD, Nelson JD, Voytek MA, Jackson GA. Ammonia-oxidizing bacterial community composition in estuarine and oceanic environments assessed using a functional gene microarray. Environ Microbiol. 2007;9:2522–2538. http://dx.doi.org/10.1111/j.1462-2920.2007.01371.x
23. Ward BB. Phytoplankton community composition and gene expression of functional genes involved in carbon and nitrogen assimilation. J Phycol. 2008;44:1490–1503. http://dx.doi.org/10.1111/j.1529-8817.2008.00594.x
24. Ward BB, Rees AP, Somerfield PJ, Joint IR. Linking phytoplankton community composition to seasonal changes in *f*-ratio. ISME J. 2011;5:1759–1770. http://dx.doi.org/10.1038/ismej.2011.50
25. Hadwin AK, Del Rio LF, Pinto LJ, Painter M, Routledge R, Moore MM. Microbial communities in wetlands of the Athabasca oil sands: Genetic and metabolic characterization. FEMS Microbiol Ecol. 2006;55(1):68–78. http://dx.doi.org/10.1111/j.1574-6941.2005.00009.x
26. Chojniak J, Wasilkowski D, Plaza G, Mrozik A, Brigmon R. Application of Biolog Microarrays Techniques for characterization of functional diversity of microbial community in phenolic-contaminated water. Int J Environ Res. 2015;9(3):785–794.
27. Matcher GF, Dorrington RA, Henninger TO, Froneman PW. Insights into the bacterial diversity in a freshwater-deprived permanently open Eastern Cape estuary, using 16S rRNA pyrosequencing analysis. Water SA. 2011;37(3):381–390. http://dx.doi.org/10.4314/wsa.v37i3.68489
28. South African National Antarctic Programme [homepage on the Internet]. No date [cited 2016 Sep 06]. Available from: http://www.sanap.ac.za/
29. South African Department of Environmental Affairs. Operation Phakisa – Oceans Economy [homepage on the Internet]. No date [cited 2016 Sep 06]. Available from: https://www.environment.gov.za/projectsprogrammes/operationphakisa/oceanseconomy. http://dx.doi.org/10.4102/sajs.v108i1/2.411

A snow forecasting decision tree for significant snowfall over the interior of South Africa

AUTHORS:
Jan Hendrik Stander[1,2]
Liesl Dyson[2]
Christien J. Engelbrecht[3]

AFFILIATIONS:
[1]South African Weather Service, Pretoria, South Africa
[2]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa
[3]Agricultural Research Council – Institute for Soil, Climate and Water, Pretoria, South Africa

CORRESPONDENCE TO:
Jan Hendrik Stander

EMAIL:
jannie.stander@weathersa.co.za

DATES:
Received: 08 June 2015
Revised: 05 Apr. 2016
Accepted: 17 Apr. 2016

KEYWORDS:
ridging surface high; geopotential thickness; cut-off low; deposition; relative humidity

HOW TO CITE:
Stander JH, Dyson L, Engelbrecht CJ. A snow forecasting decision tree for significant snowfall over the interior of South Africa. S Afr J Sci. 2016;112(9/10), Art. #2015-0221, 10 pages. http://dx.doi.org/10.17159/sajs.2016/20150221

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
South African Weather Service; University of Pretoria

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

Snowfall occurs every winter over the mountains of South Africa but is rare over the highly populated metropolises over the interior of South Africa. When snowfall does occur over highly populated areas, it causes widespread disruption to infrastructure and even loss of life. Because of the rarity of snow over the interior of South Africa, inexperienced weather forecasters often miss these events. We propose a five-step snow forecasting decision tree in which all five criteria must be met to forecast snowfall. The decision tree comprises physical attributes that are necessary for snowfall to occur. The first step recognises the synoptic circulation patterns associated with snow and the second step detects whether precipitation is likely in an area. The remaining steps all deal with identifying the presence of a snowflake in a cloud and determining that the snowflake will not melt on the way to the ground. The decision tree is especially useful to forecast the very rare snow events that develop from relatively dry and warmer surface conditions. We propose operational implementation of the decision tree in the weather forecasting offices of South Africa, as it is foreseen that this approach could significantly contribute to accurately forecasting snow over the interior of South Africa.

Significance:

- A method for forecasting disruptive snowfall is provided. It is envisaged that this method will contribute to the improved forecasting of these severe weather events over South Africa.
- Weather systems responsible for snowfall are documented and the cloud microphysical aspects important for the growth and melting of a snowflake are discussed.
- Forecasting methods are proposed for the very rare events when snow occurs over the interior of South Africa when the air is relatively dry and somewhat warmer.

Introduction

Snowfall sparks a particular interest in South Africa, especially when it descends from the mountain peaks to lower elevations. Snowfall events are well publicised by the media, and enthusiasts drive great distances to see the snow on the ground. South Africa enjoys a temperate climate, and as a result snow is not a frequent event in winter. Weather forecasters in South Africa acknowledge that they are not familiar with forecasting snowfall on lower elevations, away from the mountains.[1,2] On 7 August 2012, snow fell in the cities of Johannesburg and Pretoria in the Gauteng Province (Figure 1). Snowfall is unusual in Gauteng and is exceptionally rare in Pretoria (Table 1).[2,3] This snowfall was the first in 44 years in Pretoria, but was not forecast. Weather forecasters from the South African Weather Service (SAWS) did foresee very cold conditions and the possibility of showers of rain, but they did not anticipate the severity of the event.[2]

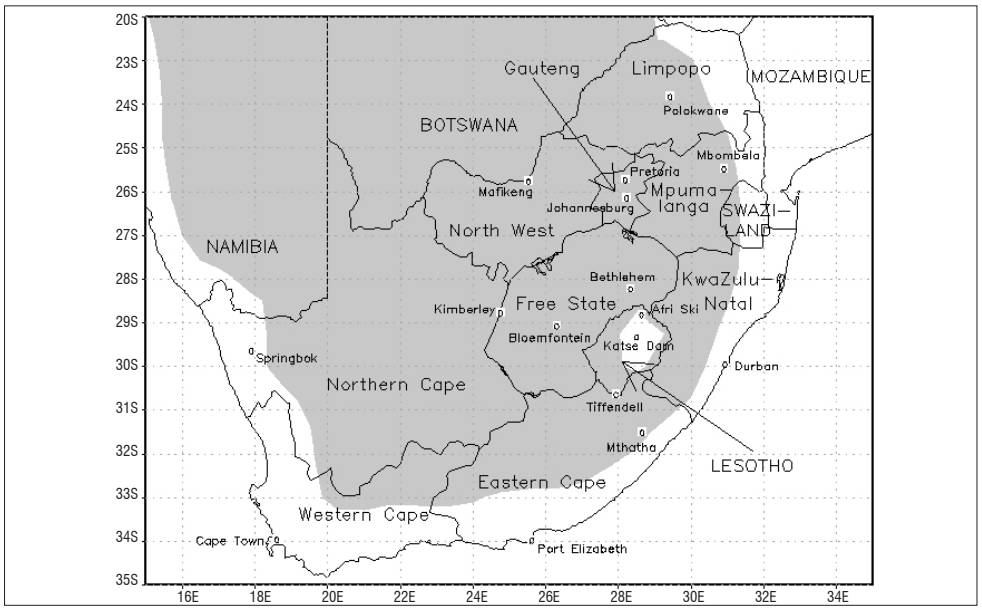


Figure 1: Location and geographical map of South Africa. The shaded area indicates where the height above mean sea level is less than 2000 m over the interior of southern Africa. Significant snowfall is defined as snowfall in this area.

Table 1: Dates of notable historical snowfall events over Johannesburg and Pretoria between 1909 and 2012[2,3]

Johannesburg	Pretoria
17 August 1909	
18 July 1915	
9 September 1921	
11 July 1926	
11 September 1936	
27 August 1962	
3 July 1963	3 July 1963 (Lyttelton)
18–19 June 1964	18 June 1964
14 July 1965	
18 October 1965	
11–12 June 1968	12 June 1968
10 September 1981	
2 July 1982	
21 July 1987	
28 June 1994	
2 August 2006	
27 June 2007	
7 August 2012	7 August 2012

Heavy snowfall is usually defined as that exceeding a threshold value. Threshold values are climate dependent and differ meaningfully by geographical and climatic region. For instance, in central Europe, heavy snowfall days are those on which the snow depth increases by 5 cm or more,[4] whereas in the mountains of Montana in the USA, heavy snowfall days are defined by 32.8 cm or more of snow.[5] Snowfall in South Africa is rare; and although snow is reported by SAWS when it occurs, the depth of the snow is not measured. An alternative definition of heavy or significant snowfall is therefore needed; in this paper, significant snowfall is defined as any amount of snowfall that occurs on the ground over the interior of South Africa in areas at an altitude of less than 2000 m above mean sea level (AMSL) (Figure 1). South Africa is characterised by an elevated plateau, rising to over 1500 m AMSL over extensive areas. The main escarpment of South Africa rises to 2000 m and higher over KwaZulu-Natal and the interior of the Eastern Cape, and in Lesotho to above 3000 m (Figure 1).[6] Snowfall is not uncommon during winter in the sparsely populated mountainous regions in South Africa; however, if snowfall occurs at lower elevations, it could be quite disruptive for the highly populated metropolises over the interior of South Africa.

There have been noteworthy snowfall events and associated negative effects during the past two decades. In July 1996, 17 people died and damage to the value of millions of rand was caused as a result of snow. Widespread communication and power cuts occurred in KwaZulu-Natal and the Free State, while major routes between KwaZulu-Natal, the Free State and Gauteng became impassable.[7] On 22 July 2002, the very cold conditions associated with snowstorms resulted in 22 lives lost, livestock loss and infrastructure destroyed in KwaZulu-Natal.[8] On the same day, many areas in the Eastern Cape had to be declared disaster areas. In June 2007, very cold conditions caused numerous power failures in KwaZulu-Natal and several roads were closed to traffic. On 27 June 2007, flights from OR Tambo International Airport (Johannesburg) were delayed by 3 h in the morning because of snow on the wings of aircrafts. Snow can cause structural icing on an aircraft and could potentially affect the aerodynamic properties, performance and weight of the aircraft. Snow can also seriously reduce horizontal visibility at an airport, making it difficult or even impossible for aircraft to land without the correct instrumentation.[9] Because of the rarity of snow at this airport, systems to deal with the snow were not in place. In September 2008, roads in KwaZulu-Natal and the Eastern Cape were closed owing to heavy snowfalls after several motor vehicle accidents occurred.[10] August 2012 was the first time on record that snow occurred simultaneously in all nine provinces of South Africa.[2] On 6 and 7 August 2012, heavy snow occurred over many parts of the country with lighter falls as far north as Pretoria.

Snow also has some positive effects on the economy of South Africa. The melting of snow in the Drakensberg Mountains has an effect on the total water budget of the area, although the exact amount is not known.[11] The groundwater level in the Table Mountain group aquifer is enhanced by snowmelt.[12] A large percentage of water in the Katse Dam comprises snowmelt that originates in the Lesotho Highlands. Water from the Katse Dam fulfils the growing demand for water in the Vaal region of South Africa.[13] Tourism also benefits from snowfall – Tiffendell Ski Resort in the Eastern Cape and Afriski Ski Mountain Resort in Lesotho require snow to remain economically viable.[8]

Between 1981 and 2011, 60 significant snow events were identified over South Africa.[2] The frequency of significant snowfall events showed strong seasonality (Figure 2), with significant snowfall occurring from May to October. These events occurred most frequently in July (>30%) followed by June (25%) and August (15–20%). Less than 15% of these events occurred in September and less than 10% in May. Although not frequently, some events (<5%) occurred in October.

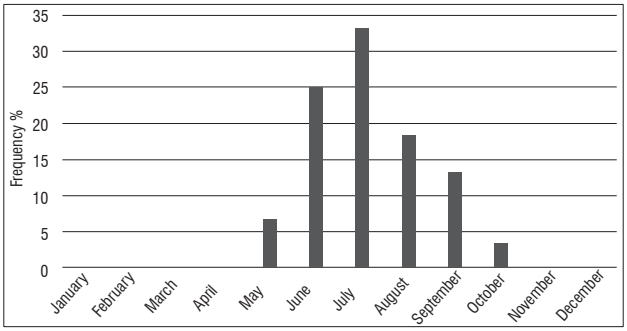


Figure 2: Frequency per month of 60 significant snow cases for the period 1981–2011.[2]

We aim to contribute to the forecasting of snowfall over South Africa by developing an operational snow forecasting decision tree (SFDT). The SFDT is especially useful to predict significant snowfall that occurs infrequently over South Africa but causes widespread disruption when it does occur. The SFDT consists of five steps or criteria that should all be met to reach a forecast of snow. The first two steps require subjective interpretation of synoptic circulation patterns and conditions conducive to the formation of precipitation. Steps 3–5 deal with ensuring that a snow flake is present in a cloud and will reach the ground without melting.

Snow forecasting decision tree

Various operational techniques or methods are used to guide forecasters in predicting the weather. Forecasters often use meteorological ingredients or decision trees to anticipate weather events. Inexperienced weather forecasters, in particular, benefit from a set of criteria or ingredients to follow in predicting rare severe events. Some examples include the ingredients that were identified to predict flash flooding.[14] Lift, instability and moisture were identified as the three main ingredients for convection.[15] A decision tree, together with threshold values, was developed to assess the potential and severity of thunderstorms.[16] Decision trees have also been used to aid in the forecasting of fog.[17,18] An ingredients-based winter season precipitation forecasting technique has been developed[19], as has a method proposed to distinguish between

freezing rain, ice pellets, rain and snow.[20] The decision tree method was also used to determine the likelihood of lake effect snowfall over Lakes Erie and Ontario.[21]

The ingredients-based approach may be limiting as it is not specific about how much of each ingredient is needed to cause a certain weather outcome.[22] The process-based approach in which fundamental physical elements are required for the formation of a certain type of weather is preferred here over the ingredients-based approach. Forecasting should not only be a recipe of ingredients, but also a list of relevant, manageable parameters on which a forecaster can focus. A forecasting method should apply logical reasoning to atmospheric physics; and several case studies should be used to evaluate the method.[23]

The SFDT uses the basic physical attributes of snow formation to guide users through a five-step process in which all the steps should be met before snow is forecast. A logical flow of the key physical mechanisms that cause significant snowfall is used to create a chronological five-step SFDT. The first two steps require subjective input and some experience and expertise on the part of the user. The last three steps deal with distinguishing between solid and liquid precipitation (Table 2).

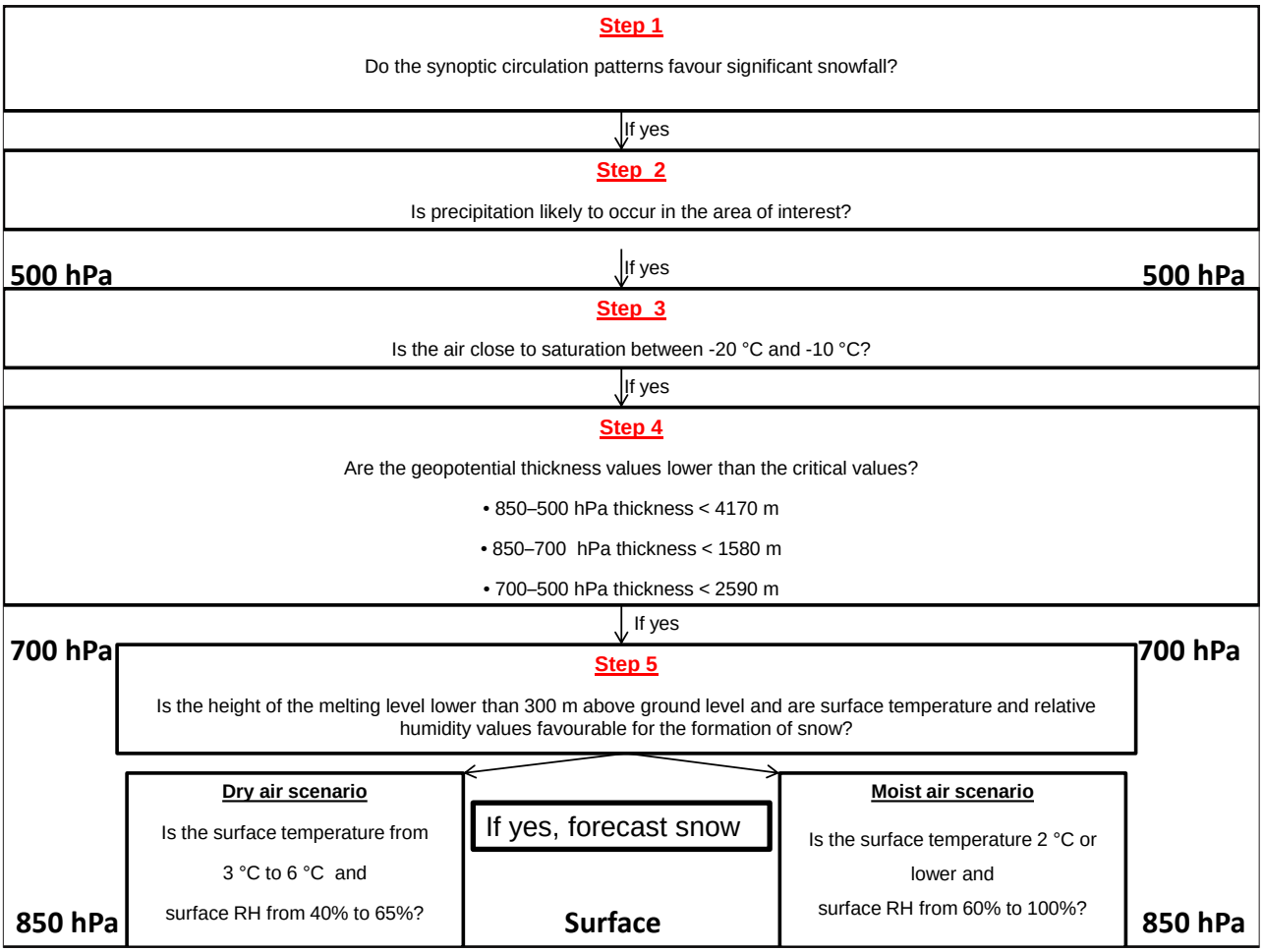
Step 1: Do the synoptic circulation patterns favour significant snowfall?

There is a rich heritage of South African synoptic climatologies and discussions of weather systems associated with rainfall.[4,24-28] Westerly wind disturbances associated with ridging anti-cyclones are important synoptic scale weather systems in winter over South Africa.[27] When an area of surface high pressure intrudes along its west–east axis into an area of lower pressure, it is said to ridge.[29] The winter rain-producing weather systems over South Africa are cold frontal troughs, low pressure systems close to land, cut-off lows (COLs) and long wave troughs.[26] Heavy precipitation in winter occurs almost exclusively from westerly wind troughs and COLs and nearly all winter rainfall in summer rainfall areas is caused by COLs.[25]

A COL is a cold-core low pressure system that is displaced equatorward (cut-off) from the westerly current. A closed low is formed in the upper troposphere and this circulation eventually extends to the surface.[25,28] COLs are weather systems of the subtropics and in the southern hemisphere are known to occur in Argentina, Uruguay, South Africa, Australia and New Zealand. An important ingredient in the development of COLs is surface cold air advection. In the South African region this cold air advection is often caused by the ridging surface high pressure system south of the land leading to cold air penetrating over the escarpment areas of the Eastern Cape and KwaZulu-Natal.[25]

Cold fronts are well defined in winter over the subtropical region of South Africa and may reach as far north as 15°S[24,30] when they are associated with upper air westerly troughs or COLs[27]. Cold fronts are normally followed by high pressure systems originating over the Atlantic Ocean. Atlantic Ocean highs (AOH) and the position and strength of these highs regularly determines the severity of the cold front. When the centre of the high is located south of the continent, the flow to its east is predominantly from south to north. The winds cross the isotherms nearly perpendicularly, causing strong cold air advection into South Africa. For the air temperature to drop by 5–10 °C over South Africa, the air is fetched by the AOH from as far south as 40–55°S.[26] It is the surface high pressure system following the cold front that sets up the horizontal pressure gradient necessary to drive cold air northwards over South Africa.

Table 2: The five-step snow forecasting decision tree



The AOH and its counterpart over the Indian Ocean, the Indian Ocean highs (IOH), are generally associated with stable weather conditions (sinking air).[29] However, when they are associated with surface cold fronts and upper air westerly troughs and COLs they are responsible for the inflow of moist unstable air into the country that may lead to widespread precipitation. Under these circumstances, conditions for precipitation are further improved over the eastern escarpment as the moisture-laden air is forced to rise because of orographic forcing.[27]

Despite this body of evidence, little work has been done in South Africa to identify the synoptic circulation patterns associated with snow. Only in a recent study was a thorough synoptic climatology constructed for snow over South Africa.[2] COLs and cold fronts, as well as surface cold air advection, have been identified in case studies[31,32] as the main synoptic drivers associated with snow in South Africa. Globally, similar weather systems have been identified as important snow-producing systems. Steep upper air troughs and closed 500-hPa lows were identified as snow-producing weather systems on the east coast of the USA, in Andorra and in Tasmania.[33-35] In North Carolina in the USA, surface anticyclones were identified as the most frequent weather system causing cold air advection necessary for snowfall to occur.[36]

In recent years, self-organising maps have been used widely in South Africa to create climatologies and to associate weather events with certain synoptic states.[18,37-39] The synoptic climatology of significant snowfall over South Africa was created using 60 snow events (426 numerical model time steps) between 1981 and 2011.[2] The 24-node self-organising map provides archetypal synoptic states associated with snow for three variables: mean sea level pressure, 500-hPa geopotential heights and 850-hPa air temperatures (Figure 3). The map places nodes with similar characteristics adjacent to each other while nodes which are very different are placed far apart.

The nodes in the first two columns in Figure 3 show a cold frontal trough southeast of the country with a 500-hPa westerly trough situated slightly west of the positions of the surface low. In all of these examples, there is evidence of the AOH ridging behind the front causing a perpendicular onshore flow onto the south and/or east coast producing very cold conditions over the country with tight temperature gradients (dotted contours in Figure 3). In the nodes in the last two columns the IOH is dominant with a southeasterly onshore flow onto the east coast where at 500 hPa a COL is situated. The nodes in the two centre columns depict the transition from the dominant cold frontal trough to the IOH. Nodes 5, 6, 11 and 12 show a low pressure situated east of the country with diminished onshore flow and with weaker temperature gradients over the southern interior.

The advantage of using the self-organising map output is that it provides a wide range of synoptic patterns all associated with snow. Figure 3 shows that there are many different surface circulation patterns associated with snowfall, but in all instances an upper air COL, or at least a sharp trough, is present. Cold fronts are important contributors to the formation of snowfall but it is the location and strength of the surface anticyclones that cause the cold air necessary for snowfall to invade South Africa.

Therefore, the first step in the SFDT (Table 1) is to compare the synoptic circulation of a particular day with those provided in Figure 3 to determine if it is appropriate to proceed to the next step in the SFDT.

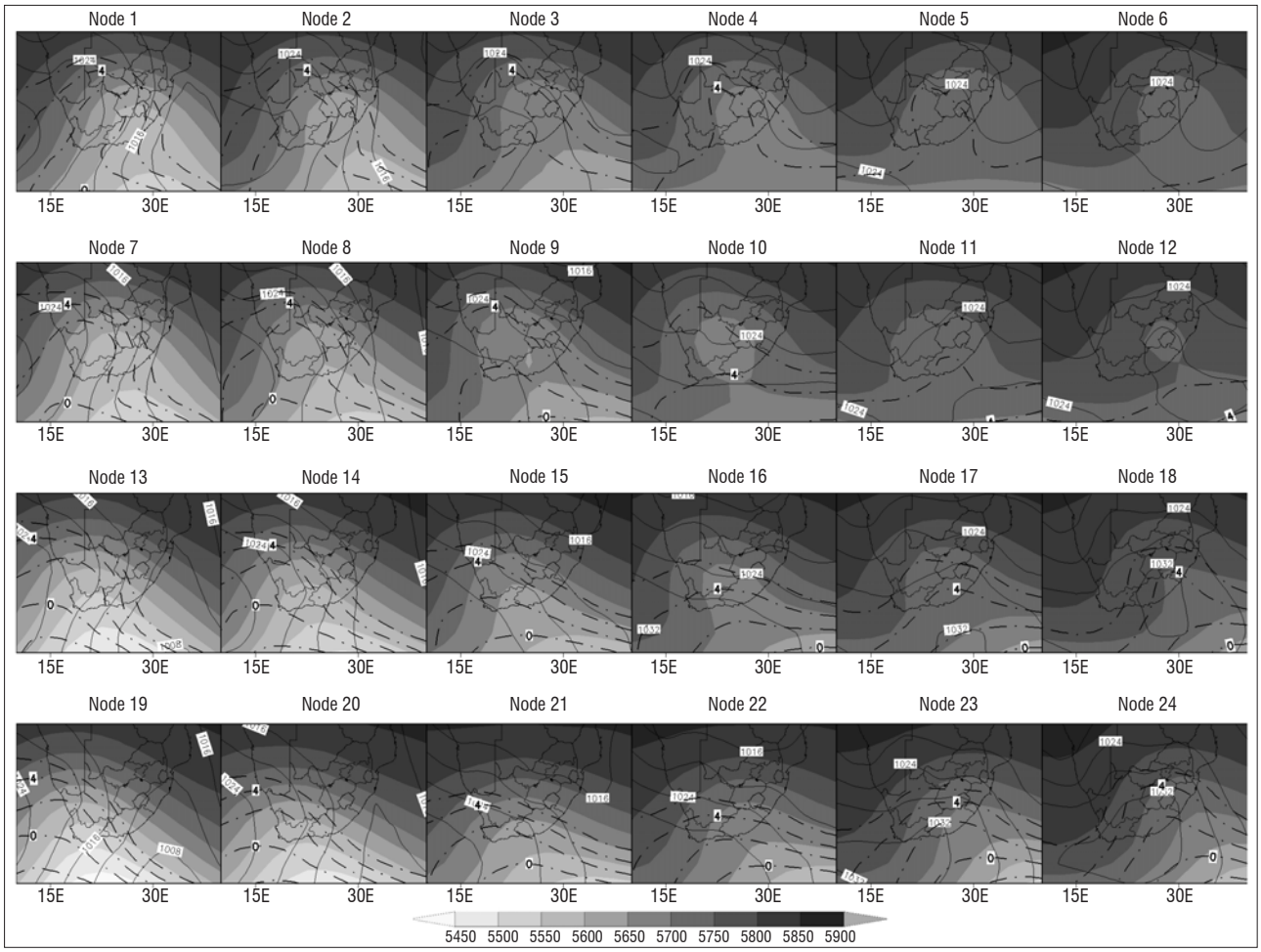


Figure 3: Archetypal self-organising maps showing sea level pressure (solid contours), 500-hPa geopotential heights (shaded) and 850-hPa temperatures lower than 6 °C (dotted/dashed contours) for 60 snow events between 1981 and 2011.

Step 2: Is precipitation likely to occur in the area of interest?

In Step 2, the user needs to determine the likelihood of precipitation using sound forecasting principles. Forecasting precipitation remains one of the major challenges for an operational meteorologist. In recent times there have been tremendous advances in the ability of numerical weather prediction products to accurately identify areas of precipitation.[40-43] However, operational weather forecasters do not exclusively rely on numerical precipitation prognoses to identify areas of precipitation – they utilise a wide variety of variables and techniques. Classical forecasting theory was developed during World War II[43] and further developed during the second part of the 20th century[14-19,20,22,44]. It is outside the scope of this paper to discuss forecasting techniques in detail, suffice to say variables such as wind divergence, vertical velocity, geopotential heights and relative humidity (RH) are investigated on several pressure levels to diagnose current and future states of the atmosphere to predict precipitation.

Once it has been determined that precipitation is likely, the user may proceed to distinguish between solid and liquid precipitation.

Step 3: Is the air close to saturation between -20 °C and -10 °C?

In the free atmosphere, water does not necessarily freeze at 0 ºC but can remain in liquid form at temperatures as low as -40 °C in the absence of nucleation nuclei.[45] Liquid water at temperatures less than 0 °C is referred to as supercooled water. Ice crystals may form and grow as a result of deposition, aggregation and riming within cold clouds (clouds with cloud top temperatures <0 °C).[45]

Deposition is the process whereby snow crystals grow by diffusion of water vapour in an environment in which the air is saturated with ice crystals. This process is also referred to as the Bergeron–Findeisen process.[45] In essence, this process takes place when water vapour changes phase to ice and the ice grows quickly when the air is saturated. Deposition normally occurs in the mid-atmosphere (700–500 hPa)[19] and where temperatures are between -16 °C and -12 °C.[45] Dendrite crystals are most likely to grow under these circumstances.[46] Dendrite ice crystals are crystals with open spaces between them that effectively cause water vapour to condense on them and to grow by deposition.[46]

When ice crystals collide and stick together to form a snowflake it is called aggregation (clumping). The efficiency of the aggregation process is greatest when the air is saturated and temperatures are between -10 °C and 0 °C.[19,47]

Riming or accretion occurs when water freezes onto ice. This process often happens when ice particles fall through supercooled droplets; the more ice crystals that are available for seeding into the low level cloud, the better the chance of producing snowflakes with larger diameters. The process is also referred to as the seeder feeder mechanism and typically occurs when ice-bearing clouds with cloud top temperatures of -15 °C (seeder clouds) move over warmer clouds containing supercooled water droplets and with cloud top temperatures of about -6 °C (feeder clouds). When the ice particles in the seeder clouds are large enough, they fall through the supercooled droplets of the feeder clouds and grow, provided the distance between the two cloud types is 1500 m or less.[19,47] If lower clouds such as stratocumulus are present, these crystals can grow by droplet riming and ice crystal aggregation as they fall through the lower layers.[47] In the absence of feeder clouds there is little chance of the ice crystals riming and forming snowflakes. The presence of low level feeder clouds increases the precipitation efficiency. In cases in which seeder clouds are absent, and only feeder clouds such as cumulus are present, light snow might be observed. However, when thick convective feeder clouds are present, heavy snowfall might occur.[47]

Cloud top temperatures of less than -10 °C indicate the presence of ice crystals in the cloud.[22,19] If the cloud top temperatures are between -15 °C and -12 °C, there is a 70–90% likelihood of ice in the cloud, while at -20 °C ice is guaranteed in the cloud. Snow can also occur with cloud top temperatures of between -10 °C and -5 °C; in these instances ice crystals grow by aggregation.[47]

In Step 3, the vertical profile of temperature and moisture should be scrutinised to establish whether ice clouds exist. If the atmosphere is close to saturation between -20 °C and -10 °C, the likelihood of fast-growing dendrite crystals through deposition is very high. In addition, saturation of the atmosphere between -10 °C and 0 °C is needed for the growth of snowflakes through aggregation.

Step 4: Are the geopotential thickness values lower than the critical values?

The SFDT needs to make provision for determining whether the snowflakes will make it all the way down to the ground without melting. This determination is done by considering the geopotential thickness for several layers in the atmosphere. The geopotential thickness is defined as the difference in geopotential heights of two pressure levels and it is proportional to the average column temperature in that layer.[44]

If the first three steps of the SFDT have been met (Table 2), the likelihood that ice crystals will be fast growing into snowflakes is high. When the snowflakes become heavy enough, they will fall to the ground under the influence of gravity. Snow will reach the ground as long as the temperature through which it falls is at or below freezing.[48] The purpose of determining geopotential thickness is to evaluate whether the temperature of the layer through which the snowflakes are falling is less than 0 °C and to identify the presence of thin warm layers that may cause the snowflakes to melt.[49]

Using thickness to identify precipitation type is a method employed widely, although threshold values vary by geographical location. In the UK the transition from rain to snow occurs when the 1000–500-hPa thickness is about 5310 m and the 1000–700-hPa thickness is 2788 m.[50] In the western USA, heavy snowfall is associated with 1000–500-hPa geopotential thickness values of 5340–5460 m.[51] Cold outbreaks with snow in Tasmania were associated with 1000–500-hPa thickness values of 5320 m.[52] Because of the height AMSL of the interior of South Africa, a pressure level close to the ground surface must be used for the higher pressure threshold in the calculation of thickness values. Surface pressures over the interior plateau are close to 850 hPa[6] and this pressure level is used as the surface pressure threshold. The 850–500-hPa geopotential thickness of less than 4170 m was found to occur in conjunction of significant snowfall over the interior of South Africa.[2]

Examining the values of thinner geopotential thicknesses or partial thicknesses is important to ensure that there are no thin warm layers in the atmosphere that could cause the snowflakes to melt. The partial thickness in the layer between the ground and the freezing level is especially significant as it is indicative of cold air close to the ground in support of frozen precipitation. An 850–700-hPa partial thickness value of less than 1552 m implies that the average column temperature in that layer is less than 0 °C and therefore favourable for snow. In a New York study, the thickness of the same level was less than 1550 m for snow to occur.[49] Over the interior of South Africa, significant snowfall was associated with 850–700-hPa thickness values of less than 1580 m.[2]

When the average column temperature in the 700–500 hPa layer is 0 °C, the thickness is 2690 m. However, during the snow event over Bloemfontein and Johannesburg in June 1964, a 700–500-hPa thickness map was analysed from upper air sounding data at Bloemfontein, Irene and Durban. The thicknesses in this instance were 100 m less than the 2690 m threshold and varied between 2520 m and 2560 m.[53] Significant snowfall over the interior of South Africa was found to occur when the 500–700-hPa thickness values were less than 2590 m.[2]

Step 4 of the SFDT requires that the geopotential thickness for the 850–500-hPa layer, as well as that for the 700–500-hPa and 850–700-hPa layers, needs to be less than the critical values in Table 2.

Step 5: Are the melting level height, surface temperature and relative humidity values suitable for snow?

Snow is more likely to reach the ground when the melting level is not more than 300 m above ground level (AGL).[45,54] The melting level is the height of the 0 °C isotherm AGL. Snow can still reach the ground when

the temperature between the ground and the freezing level is greater than 0 °C. This happens when snowflakes start to melt once they fall through the melting level into the dryer air close to the ground. Melting leads to the absorption of latent heat that cools surrounding temperatures, allowing some flakes to make it to the ground.[36] The absorption of latent heat by melting could also cause the air surrounding the snowflake to be cooled to freezing. The snowflake can fall several hundreds of metres below the melting level before melting completely when the environmental temperature is 5 °C. When surface RH is less than 90% and the temperature lapse rate is wet adiabatic, the snowflakes can fall approximately 600 m in above freezing temperatures before melting. As the RH increases, the distance that a snowflake can fall before melting becomes shorter.[45]

The RH in the lower levels of the atmosphere is a determining factor in the surface temperature required for snow to reach the ground. When surface RH is close to 100%, snow occurs in temperatures below 1 °C and rainfall is more likely than snow when temperatures are greater than 1 °C. For surface RH values less than 90%, snow occurs regularly when surface temperatures are lower than 2.5 °C.[55] Snow occurs at temperatures lower than 2 °C when surface RH exceeds 60%, because in high RH environments, the water vapour density is higher than that of the snowflake and condensation of water takes place onto the snowflake. This release of latent heat melts the snowflake. When surface RH values are less than 60%, surface temperatures must be between 4 °C and 6 °C for snow to occur.[45] In environments with low RH, the water vapour density is less than the snowflake and sublimation of water vapour occurs from the snowflake, cooling the flake and the subsequent ambient air temperature in that location.[45,55]

In Step 5, two snowfall scenarios are defined. The moist air scenario and the dry air scenario. When the surface temperature is 2 °C or lower and surface RH is between 60% and 100%, the moist air scenario occurs. In moist air scenarios, the air is saturated close to the surface with the presence of stratiform and nimbostratus clouds that extend from close to the surface to mid-levels. The dry air scenario occurs when the surface is dry (RH<65%) and surface temperatures are 3–6 °C. (Table 2). In dry air scenarios, the atmosphere near the surface is dry and clouds which can form in low RH environments should be present, such as cold air cumulus[56] (convective) clouds which develop behind cold fronts.[57] The user of the SFDT has to determine whether the snow will melt. Melting typically occurs when cold air advection and upward motion in the lower levels are fairly weak.[49,58] Furthermore, in areas in which it has been raining for a while, cooling from melting will be strongest and the absorption of latent heat in these areas further cools the atmosphere, which can lead to snow.[58]

Case study: 7 August 2012, Gauteng

On 6 and 7 August 2012, snowfall occurred in all nine provinces of South Africa. On 7 August 2012, widespread snowfall occurred over nearly the entire Gauteng Province. Virtually all of the aerodromes over the province reported snowfall, including OR Tambo International, Rand, Lanseria and Wonderboom Airports. Reports of snow came from all the metropolises in the province. Between 1909 and 2012, snowfall has been reported on only three occasions in both Pretoria and Johannesburg (Table 1). Prior to 7 August 2012, the last report of snowfall in Pretoria was 12 June 1968. One of the reasons that makes the snowfall of 7 August 2012 over Gauteng exceptional, is that it occurred in relatively warmer yet distinctively drier surface conditions. It is shown that by applying the SFDT, this very rare snow event could have been anticipated. The snow that occurred in Bloemfontein on 7 August 2012 is also analysed using the SFDT, as it is an example of the moist snow scenario. The two contrasting conditions are used to demonstrate the effectiveness of the SFDT.

Data used in the case study

The synoptic circulation patterns on 7 August 2012 are illustrated using National Centers for Environmental Prediction reanalysis two (NCEP 2) data.[59] Vertical profiles of temperature, dew point temperature and RH at Irene (situated between Pretoria and Johannesburg) and Bloemfontein were obtained from SAWS upper air ascents. The geopotential thicknesses and melting level heights at Irene and Bloemfontein were also calculated from these data. Surface temperature and RH values were measured by SAWS automatic weather stations.

Meteosat Second Generation satellite data from EUMETSAT were used to determine the properties of the clouds.[60] The infrared (IR) 10.8-μm channel is useful to determine cloud top temperatures. High reflectivity of clouds on the high-resolution visual images indicates the presence of ice in clouds. Furthermore the cloud structure of clouds in the visual image are used to distinguish between convective and layered cloud. Convective cloud has a clear cellular pattern whereas layered cloud appear as uniformly grey sheets of cloud.[60]

Application of the snow forecasting decision tree

Step 1: Do the synoptic circulation patterns favour significant snowfall?

Figure 4 depicts the synoptic circulation pattern on 7 August 2012. The surface cold front was located east of the country with the 500-hPa COL over the eastern interior of South Africa. There was a strong onshore flow of moisture onto the south and southeast coasts and an AOH ridging at 40 °S behind the cold front. This circulation pattern is very favourable for significant snowfall and represents the synoptic patterns depicted in the top right-hand corner of Figure 3. Although surface temperatures were quite cold, there was a weak temperature gradient over the eastern half of the country. Step 1 of the SFDT is therefore met.

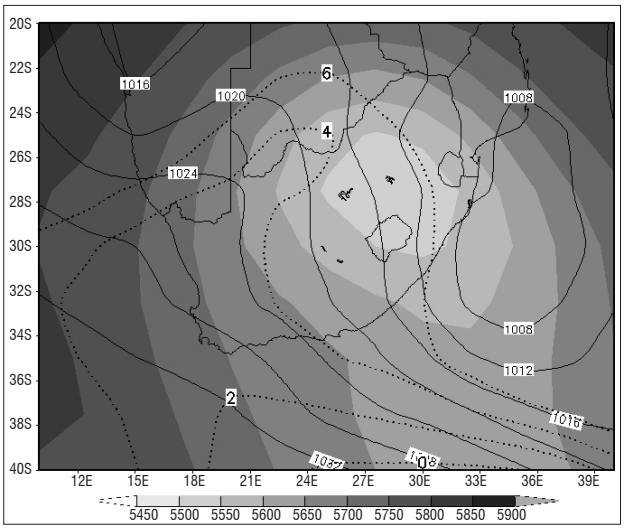


Figure 4: Sea level pressure (solid contours), 500-hPa geopotential heights (shaded) and 850-hPa temperatures lower than 6 °C (dotted contours) on 7 August 2012 at 1200 UTC.

Step 2: Is precipitation likely to occur in the area of interest?

The cyclonic circulation of the 850-hPa winds indicate that the 850-hPa low was situated over the coast of KwaZulu-Natal, causing large areas of surface convergence (negative divergence) over the eastern parts of the country – including Gauteng and the eastern Free State (Figure 5). The cyclonic circulation around the low caused a strong onshore flow of moisture over the southeast and east coasts, resulting in RH values in excess of 80% over the southern half of South Africa but with drier conditions over Gauteng. Over the Free State, the RH values were 50–70%. At 500 hPa, the low was situated over the eastern interior (Figure 4) with areas of upper air divergence over the eastern half of the country (not shown). Conditions for precipitation were therefore quite favourable over Bloemfontein and Gauteng.

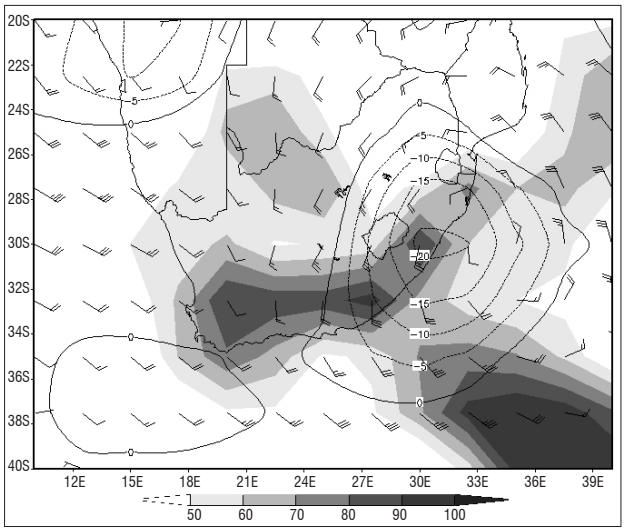


Figure 5: The 850-hPa winds (knots), relative humidity (%, shaded) and horizontal wind convergence (x10^5 s^{-1}, contour values) on 7 August 2012 at 1200 UTC.

Step 3: Is the air close to saturation between -15 °C and -12 °C?

Data from Irene upper air sounding at 1200 UTC on 7 August 2012 is depicted in Figure 6. The surface temperature was 6 ºC with RH values less than 50% in the surface layers. In those layers in the atmosphere (600–700 hPa), at which temperatures were between -20 °C and -10 °C, RH values varied from 54% to 60%. Conditions were therefore favourable for fast-growing dendrite crystals through deposition.

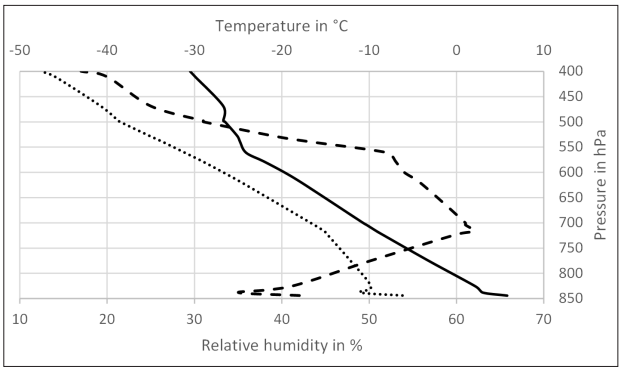
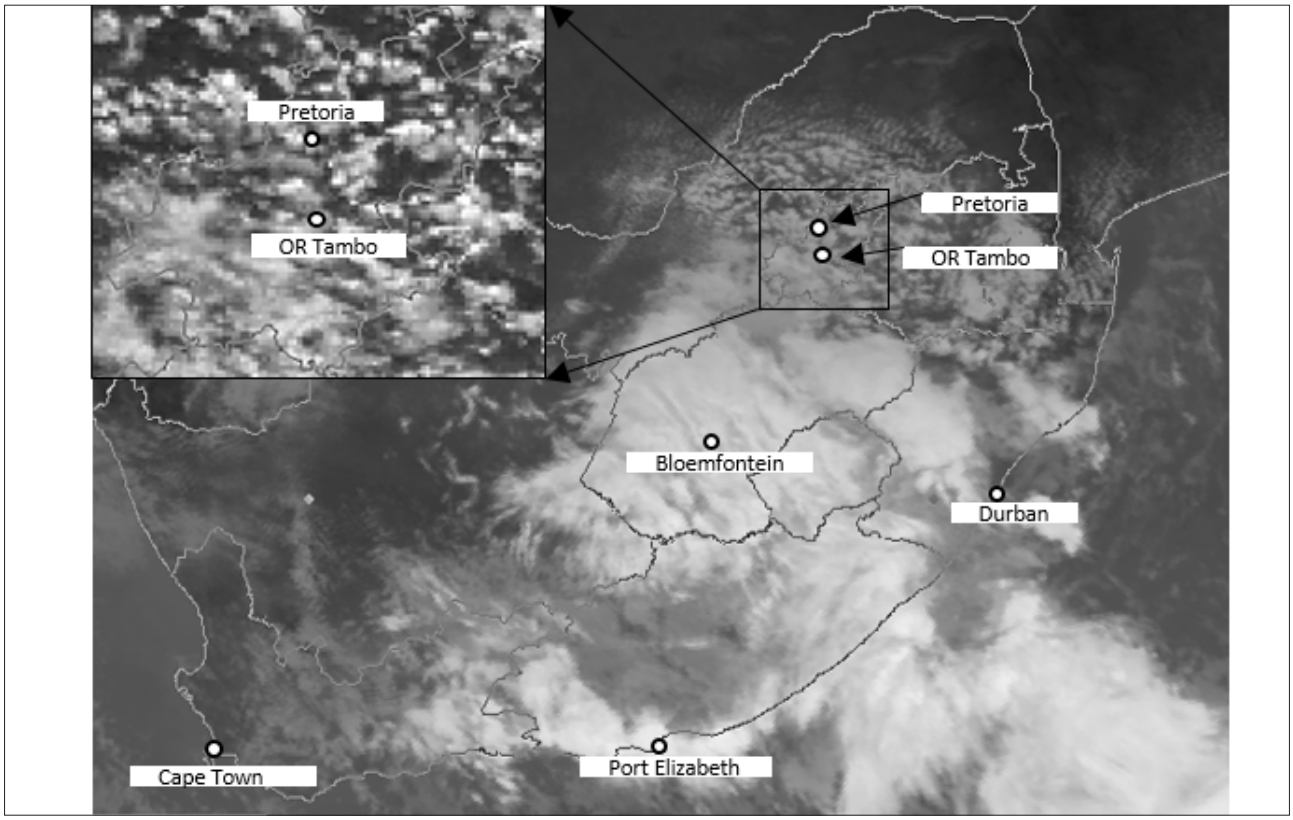


Figure 6: The vertical profile of temperature (solid), dew point temperature (dotted) and relative humidity (dashed) as obtained from upper air ascent at Irene on 7 August 2012 at 1200 UTC.

The IR 10.8-μm Meteosat Second Generation satellite image indicates that the temperatures of the cloud tops over Gauteng were between -10 °C and -20 °C (Figure 7). Cloud top temperatures in this range indicate the presence of ice in the clouds.[22] The high reflectivity of these clouds (bright white colours) is further evidence of ice in the clouds. The convective nature of the clouds is illustrated by the cellular structure of the clouds over Gauteng (Figure 7 inset). The satellite image shows that there were no higher (seeder) clouds present over Gauteng and this is also reflected in the very dry conditions at pressure levels lower than 550 hPa (Figure 6). Over Gauteng, ice crystals primarily formed by deposition and aggregation were not dominant in this instance.

Over Bloemfontein the atmosphere was saturated throughout, with surface temperatures lower than 2 °C and RH values greater than 90%



©2012 Eumetsat

Figure 7: Meteosat Second Generation satellite IR 10.8-μm image over South Africa on 7 August 2012 at 1200 UTC. Inset: High-resolution visible image zoomed in over Gauteng.

(Figure 8). Considering the uniformly grey sheet of cloud in Figure 7, it is clear that seeder and feeder clouds were present over the Free State, causing ice crystals to grow by both deposition in the higher seeder clouds and aggregation in the low-level feeder clouds. Snow occurred during the late morning and early afternoon in Bloemfontein, but continued until the evening at Bethlehem.

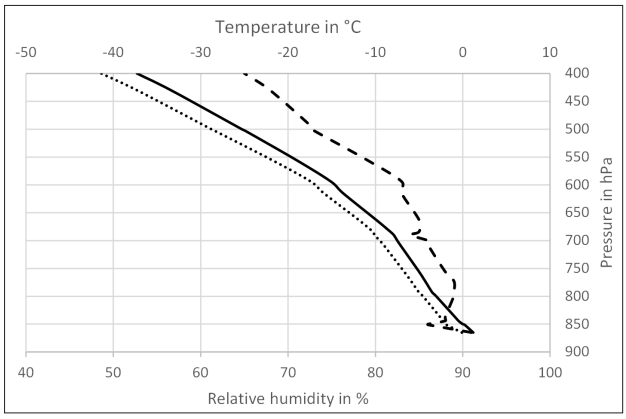


Figure 8: The vertical profile of temperature (solid), dew point temperature (dotted) and relative humidity (dashed) as obtained from upper air ascent at Bloemfontein on 7 August 2012 at 1200 UTC.

Step 4: Are the geopotential thickness values lower than the critical values?

Table 3 depicts the geopotential thickness values over Irene and Bloemfontein for the three critical levels identified by the SFDT. Over both Bloemfontein and Irene, the thickness values for all three levels were well below the critical values identified for significant snowfall over South Africa (Table 2). This indicates that the snowflakes would not have fallen through warm layers that may have caused it to melt.

Step 5: Are the melting level height, surface temperature and relative humidity values suitable for snow?

At Irene and Bloemfontein, the first four steps of the SFDT were met and the difference between the moist and dry air scenario is illustrated by considering the surface conditions at these two locations. The height of the melting level at Irene was 420 m AGL – higher than the threshold value[45,54] of 300 m (Table 2). Considering all the other favourable factors over Gauteng, this value alone is not enough to reject the possibility of snowfall. At Bloemfontein, the melting level was only 141 m AGL and therefore well within the critical value.

Over Irene the surface RH was only 42% and according to the SFDT snow will occur in the dry air scenario if surface temperatures are between 3 °C and 6 °C. Surface temperature at Irene was 5 °C (Table 3) and even lower in Pretoria and Johannesburg (not shown). Surface conditions therefore met the criteria for the dry air snow scenario. Snow occurred at 1200 UTC in Pretoria and Irene. At OR Tambo International Airport light snowfall started before midday and continued well into the evening.

At Bloemfontein the surface RH was 91%, which clearly indicates that this was a moist snow scenario. The surface temperature was only 1 °C, which meets the criterion needed for snow to occur during a moist air scenario. Snow occurred through most of the morning in Bloemfontein, with heavy snow showers reported from time to time during the day.

Discussion and conclusions

The World Meteorological Organization defines a snow day as a day with a snow depth greater than a certain threshold.[61] For instance, in Switzerland this threshold is set at 1 cm.[62] Because of the rarity of snow in South Africa, the snow depth is not measured; however, the occurrence of snow is noted at SAWS offices. The absence of snow depth measurements necessitates an alternative definition of significant snowfall over South Africa. Significant snowfall is therefore defined in this paper as snow that reaches the ground in areas at an altitude of less than 2000 m AMSL. Snowfall in South Africa normally occurs in the sparsely populated mountainous regions. At lower altitudes, the snow often melts before reaching the ground, but when it does reach the ground it causes widespread disruption to infrastructure and even loss of life. Significant snowfall events are rare and irregular. On average there are two significant snowfall events per year over South Africa[2] and these events do not occur every year. In comparison, there were 160 snow days on average at Semedan in the central Alps.[62] As significant snow occurs irregularly with only, at most, a few events per year, it becomes important to understand the synoptic conditions and atmospheric variables which cause snowfall at lower elevations in order to predict it when it does occur. A snow forecasting decision tree (SFDT) is proposed, in which the synoptic circulation patterns, cloud microphysical aspects and other atmospheric variables are condensed into a user-friendly product.

The SFDT consists of five steps that should all be met to forecast snowfall. The first step identifies the snow synoptic circulation type and the second step requires the user to determine if precipitation is likely to occur. Steps 3 to 5 distinguish between solid (snow) and liquid (rain) precipitation by first ensuring that ice crystals are present in the cloud and then investigating if the snow will melt on the way to the ground. In the last step, the user needs to differentiate between *wet* and *dry* snow conditions by examining surface temperature and RH values.

The synoptic circulation systems associated with significant snowfall over South Africa are COLs or sharp troughs at 500 hPa together with surface systems which cause strong cold air advection into the subcontinent. The surface circulation is broadly categorised into cold frontal troughs followed by the ridging AOH and the IOH southeast of the country. These weather systems are similar to those that have been identified elsewhere in the world.[33-36]

Critical geopotential thickness values associated with snow have been identified for several layers in the atmosphere (Table 2), but were specifically developed for the interior of South Africa which lies close to the 850-hPa pressure level. Comparable threshold values have been developed internationally. For instance, the UK Met Office gives a 90% chance of snow if the 1000–500-hPa thickness is less than 5180 m and the 1000–850-hPa thickness is less than 1281 m. The UK Met Office considers the melting level to be at about 108 m above the ground, whereas we identified 300 m as the threshold value.[63]

Forecasting snow under wet surface conditions is a well-known technique in weather forecasting offices in South Africa.[2] The real value of the SFDT is in aiding in the identification of the very rare dry snow conditions.

Table 3: Geopotential thickness values, melting level heights and surface temperature and relative humidity values at Irene and Bloemfontein on 7 August 2012 at 1200 UTC

	Thickness (m)			Surface		
	850–500 hPa	850–700 hPa	700–500 hPa	Melting level height (m)	Temperature (°C)	Relative humidity (%)
Irene	3967	1477	2490	420	5	42
Bloemfontein	4065	1532	2533	141	1	91

This paper contributes to a better understanding of the synoptic circulation systems associated with significant snow and adds insight into the cloud microphysical and surface conditions during snowfall. It is recommended that the SFDT is implemented in forecast offices of the SAWS, which will allow further fine-tuning of critical values. Operational weather forecasters can add value to the SFDT by providing feedback on its usefulness in an operational environment. Steps 3 to 5 could be automated and made available through numerical weather prediction output. However, subjective and human interpretation of Steps 1 and 2 remains a necessity.

Acknowledgements

We thank the South African Weather Service for making data available. The library personnel provided invaluable support in identifying archived snow events and their support is acknowledged.

Authors' contributions

This paper emanates from the work that J.H.S. conducted to obtain his MSc at the University of Pretoria under the supervision of L.D. J.H.S. conducted all the research and constructed the forecasting decision tree. L.D. provided scientific guidance, and helped with the preparation of the manuscript. C.J.E. developed the initial Fortran code used in this study and provided feedback on the manuscript.

References

1. Gambrell J. South Africa snowfall stuns Johannesburg [homepage on the Internet]. c2012 [cited 2016 Aug 4]. Available from: http://www.govtslaves.info/rare-snowfall-stuns-much-of-south-africa/
2. Stander JH. Synoptic circulation patterns and atmospheric variables associated with significant snowfall over South Africa in winter [MSc dissertation]. Pretoria: University of Pretoria; 2013.
3. Viljoen F. Besondere sneeuvalle in die geskiedenis van Suid-Afrika [Noteworthy snow events in the history of South Africa]. South African Weather Bureau Newsletter. 1988 June; p. 1–8. Afrikaans.
4. Bednorz E. Synoptic reasons for heavy snowfalls in the Polish–German lowlands. Theor Appl Climatol. 2008;9:133–140. http://dx.doi.org/10.1007/s00704-007-0322-4
5. Birkeland KW, Mock CJ. Atmospheric circulation patterns associated with heavy snowfall events, Bridger Bowl, Montana, USA. Mt Res Dev. 1996;16:281–286. http://dx.doi.org/10.2307/3673951
6. Taljaard JJ. Atmospheric circulation systems, synoptic climatology and weather phenomena of South Africa. Part 1: Controls of the weather and climate of South Africa. South African Weather Bureau Technical Paper. Pretoria: South African Weather Bureau; 1994. p. 27.
7. De Waal K. Nuus van die weerkantore [News from the weather offices]. South African Weather Bureau Newsletter. 1996 July; p. 13–17. Afrikaans.
8. Muirhead R. World weather disasters: July 2002. J Meteorol. 2002;273:359–360.
9. World Meteorological Organization. World Meteorological Organization Training Programme: Aviation hazards. ERT-20, WMO/TD-No. 1390. Geneva: World Meteorological Organization; 2007.
10. KZN sprokiesmooi onder die sneeu [KZN a fairytale under snow]. Beeld. 2008 September 22; p. 2. Afrikaans.
11. Nel W, Sumner PD. First rainfall data from the KZN Drakensberg escarpment edge (2002 and 2003). Water SA. 2005;31(3):399–402.
12. Wu Y, Xu Y. Snow impact on groundwater recharge in Table Mountain Group aquifer systems with a case study of the Kommissiekraal River catchment South Africa. Water SA. 2005;3(3):275–278.
13. Wright JS. The impact of Katse Dam water on water quality in the Ash, Liebenbergsvlei and Wilge Rivers and the Vaal Dam [MSc dissertation]. Johannesburg: University of Johannesburg; 2006.
14. Doswell CA, Brooks HE, Maddox RA. Flash flood forecasting: An ingredients-based methodology. Weather Forecast. 1996;11:560–581. http://dx.doi.org/10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2
15. Schultz DM, Schumacher PN. The use and misuse of conditional symmetric instability. Mon Weather Rev. 1999;127:2709–2732. http://dx.doi.org/10.1175/1520-0493(1999)127<2709:TUAMOC>2.0.CO;2
16. Mills GA, Colquhoun JR. Objective prediction of severe thunderstorm environments: Preliminary results linking a decision tree with an operational regional NWP model. Weather Forecast. 1998;13(4):1078–1092. http://dx.doi.org/10.1175/1520-0434(1998)013<1078:OPOSTE>2.0.CO;2
17. Tardif R, Rasmussen RM. Event-based climatology and typology of fog in the New York City region. J Appl Meteorol Climatol. 2007;46:1141–1168. http://dx.doi.org/10.1175/JAM2516.1
18. Van Schalkwyk L, Dyson L. Climatological characteristics of fog at Cape Town International Airport. Weather Forecast. 2013;28:631–646. http://dx.doi.org/10.1175/WAF-D-12-00028.1
19. Wetzel SW, Martin JE. Forecasting technique – An operational ingredients-based methodology for forecasting midlatitude winter season precipitation. Weather Forecast. 2001;16:156–167. http://dx.doi.org/10.1175/1520-0434(2001)016<0156:AOIBMF>2.0.CO;2
20. Bourgouin P. A method to determine precipitation types. Weather Forecast. 2000;15:583–592. http://dx.doi.org/10.1175/1520-0434(2000)015<0583:AMTDPT>2.0.CO;2
21. Niziol TA. Operational forecasting of lake effect snowfall in western and central New York. Weather Forecast. 1987;2:310–321. http://dx.doi.org/10.1175/1520-0434(1987)002<0310:OFOLES>2.0.CO;2
22. Schultz DM, Cortinas Jr JV, Doswell III CA. 2002: Comments on "An operational ingredients-based methodology for forecasting midlatitude winter season precipitation". Weather Forecast. 2002;17:160–167. http://dx.doi.org/10.1175/1520-0434(2002)017<0160:COAOIB>2.0.CO;2
23. Chaston P. The magic chart for forecasting snow amounts. Natl Weather Digest. 1989;14(2):20–22.
24. Karoly DJ, Vincent DG, editors. Meteorology of the southern hemisphere. Meteorol Monogr 27. Boston, MA: American Meteorological Society; 1998. http://dx.doi.org/10.1175/0065-9401-27.49.1
25. Taljaard JJ. Cut-off lows in the South African region. South African Weather Bureau Technical Paper. Pretoria: South African Weather Bureau; 1985. p. 14.
26. Taljaard JJ. Atmospheric circulation systems, synoptic climatology and weather phenomena of South Africa. Part 6: Rainfall in South Africa. South African Weather Bureau Technical Paper. Pretoria: South African Weather Bureau; 1996. p. 32.
27. Tyson PD, Preston-Whyte RA. The weather and climate of southern Africa. 2nd ed. Cape Town: Oxford University Press; 2000.
28. Singleton AT, Reason CJC. Variability in the characteristics of cut-off low pressure systems over subtropical southern Africa. Int J Climatol. 2007;27(3):295–310. http://dx.doi.org/10.1002/joc.1399
29. Van Heerden J, Hurry L. Southern Africa's weather patterns: An introductory guide. Pretoria: Via Africa Limited, Acacia Books; 1995.
30. Taljaard JJ, Schmitt W, Van Loon H. Frontal analysis with application to the southern hemisphere. South African Weather Bureau Notos 10. Pretoria: South African Weather Bureau; 1961. p. 25–58.
31. Mulder N, Grab SW. Contemporary spatio-temporal patterns of snow cover over the Drakensberg. S Afr J Sci. 2009;105(5/6):228–233.
32. De Villiers MP. Winter weather at Tiffendell. South African Weather Service internal report B0243: No FOR/20. Unpublished report; 2001.
33. Kocin PJ, Uccellini LW. Northeast snowstorms: Overview. Meteorol Monogr volume 32 no. 54. Boston, MA: American Meteorological Society; 2004. http://dx.doi.org/10.1007/978-1-878220-32-5
34. Esteban P, Jones PD, Martín-Vide J, Mases M. Atmospheric circulation patterns related to heavy snowfall days in Andorra, Pyrenees. Int J Climatol. 2004;25:319–329. http://dx.doi.org/10.1002/joc.1103
35. Jones MC. Climatology of cold outbreaks with snow over Tasmania. Australian Meteorological Magazine. 2003 July;52:157–169.
36. Cuviello MP. A model for refining precipitation-type forecasts for winter weather in the Piedmont region of North Carolina on the basis of partial thickness and synoptic weather patterns [MA thesis]. Chapel Hill, NC: University of North Carolina; 2007.

37. Tennant W, Hewitson BC. Intra-seasonal rainfall characteristics and their importance to the seasonal prediction problem. Int J Climatol. 2002;22:1033–1048. http://dx.doi.org/10.1002/joc.778

38. Engelbrecht CJ, Landman WA, Engelbrecht FA, Malherbe J. A synoptic decomposition of rainfall over the Cape south coast of South Africa. Clim Dynamics. 2015:44;2589–2607. http://dx.doi.org/10.1007/s00382-014-2230-5

39. Lennard C, Hegerl G. Relating changes in synoptic circulation to the surface rainfall response using self-organising maps. Clim Dyn. 2015;44:861–879.

40. Landman S, Engelbrecht FA, Engelbrecht CJ, Dyson LL, Landman WA. A short-range weather prediction system for South Africa based on a multi-model approach. Water SA. 2012;38(5):765–774. http://dx.doi.org/10.4314/wsa.v38i5.16

41. Bopape MJM, Engelbrecht FA, Randall DA, Landman WA. Advances towards the development of a cloud-resolving model in South Africa. S Afr J Sci. 2014;110(9–10), Art. #2013-0133, 12 pages. http://dx.doi.org/10.1590/sajs.2014/20130133

42. Rezacova D, Szintai B, Jakubiak B, Yano JI, Turner S. Verification of high resolution precipitation forecast by radar-based data. In: Plant RS, Yano J-I, editors. Parameterization of atmospheric convection. London: Imperial College Press; 2014. p. 2.

43. Sutcliffe RC. A contribution to the problem of development. QJR Meteorol Soc. 1947;73:370–383. http://dx.doi.org/10.1002/qj.49707331710

44. Holton JR. An introduction to dynamic meteorology. 5th ed. Waltham, MA: Academic Press; 2013. http://dx.doi.org/10.1016/B978-0-12-384866-6.00001-5

45. Pruppacher HR, Klett JD. Microphysics of clouds and precipitation. 2nd ed. London: Springer; 2010. http://dx.doi.org/10.1007/978-0-306-48100-0_2

46. Ahrens CD. Meteorology today: An introduction to weather, climate and the environment. Belmont, CA: Thomson Brooks/Cole; 2007.

47. Jiusto JE, Weickmann HK. Types of snowfall. Bull Am Meteorol Soc. 1973;54(11):1148–1162. http://dx.doi.org/10.1175/1520-0477(1973)054<1148:TOS>2.0.CO;2

48. Beckman SK. Use of enhanced IR/visible satellite imagery to determine heavy snow areas. Mon Weather Rev. 1987;115:2060–2087. http://dx.doi.org/10.1175/1520-0493(1987)115<2060:UOEISI>2.0.CO;2

49. Heppner POG. Snow versus rain: Looking beyond the "magic" numbers. Weather Forecast. 1992;7(4):683–691. http://dx.doi.org/10.1175/1520-0434(1992)007<0683:SVRLBT>2.0.CO;2

50. Lamb HH. Two-way relationship between the snow or ice limit and 1,000-500 mb thickness in the overlying atmosphere. QJR Meteorol Soc. 1954;81:172–189. http://dx.doi.org/10.1002/qj.49708134805

51. Younkin RJ. Circulation patterns associated with heavy snowfall over the western United States. Mon Weather Rev. 1968;96:851–853. http://dx.doi.org/10.1175/1520-0493(1968)096<0851:CPAWHS>2.0.CO;2

52. Jones MC. Climatology of cold outbreaks with snow over Tasmania. Australian Meteorological Magazine. 2003 July;52(3):157–169.

53. Stranz D, Taljaard JJ. Analysis of an abnormal winter situation in South Africa during June 1964. South African Weather Bureau Notos 14. Pretoria: South African Weather Bureau; 1964. p. 7–32.

54. Wallace JM, Hobbs PV. Atmospheric science: An introductory survey. Vol. 92. Waltham, MA: Academic Press; 2006.

55. Matsuo T, Sasyo Y. Non-melting phenomena of snowflakes observed in sub saturated air below freezing level. J Meteorol Soc Japan. 1981;59:26–32.

56. Derbyshire SH, Beau I, Bechtold P, Grandpeix JY, Piriou JM, Redelsperger JL, et al. Sensitivity of moist convection to environmental humidity. QJR Meteorol Soc. 2004;130:3055–3079. http://dx.doi.org/10.1256/qj.03.130

57. Miura Y. Aspect ratios of longitudinal rolls and convection cells observed during cold air outbreaks. J Atmos Sci. 1986;43:26–39. http://dx.doi.org/10.1175/1520-0469(1986)043<0026:AROLRA>2.0.CO;2

58. Kain JS, Goss SM, Baldwin ME. The melting effect as a factor in precipitation-type forecasting. Weather Forecast. 2000;15:700–713. http://dx.doi.org/10.1175/1520-0434(2000)015<0700:TMEAAF>2.0.CO;2

59. Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, et al. The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc. 1996;77:437–471. http://dx.doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

60. Eumetrain. Satellite image interpretation [course on the Internet]. No date [cited 2015 May 06]. Available from: http://eumetrain.org/courses/satellite_image_interpretation.html

61. World Meteorological Organization. Handbook on climate and climate temp reporting. WMO TD 1188. Geneva: World Meteorological Organization; 2009.

62. Foppa N, Seiz G. Inter-annual variations of snow days over Switzerland from 2000–2010 derived from MODIS satellite data. Cryosphere. 2012;6:331–342. http://dx.doi.org/10.5194/tc-6-331-2012

63. UK Met Office. Source book to the forecaster's reference book. Berkshire: The Met Office; 1997. Available from: http://www.metoffice.gov.uk/media/pdf/9/3/Source_Book_to_Forecasters_Reference_Book_Complete.pdf

New light on vitamin B_{12}: The adenosylcobalamin-dependent photoreceptor protein CarH

AUTHOR:
Susan M. Chemaly[1]

AFFILIATION:
[1]School of Chemistry, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Susan Chemaly

EMAIL:
susan.chemaly@wits.ac.za

DATES:
Received: 06 Apr. 2016
Revised: 07 June 2016
Accepted: 30 June 2016

KEYWORDS:
photolysis; coenzyme B_{12}; light-sensing protein; DNA transcription; X-ray crystallography

HOW TO CITE:
Chemaly SM. New light on vitamin B_{12}: The adenosylcobalamin-dependent photoreceptor protein CarH. S Afr J Sci. 2016;112(9/10), Art. #2016-0106, 9 pages. http://dx.doi.org/10.17159/sajs.2016/20160106

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

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

Adenosylcobalamin (AdoCbl), or coenzyme B_{12}, is a cofactor for enzymes important in metabolism in humans (and other mammals) and bacteria. AdoCbl contains a Co-C bond and is extremely light sensitive, but, until recently, this light sensitivity appeared to have no physiological function. Recently, AdoCbl has been found to act as cofactor for a photoreceptor protein (CarH) that controls the expression of DNA coding for transcription of the proteins needed for synthesis of carotenes in certain non-photosynthetic bacteria. In 2015, the X-ray crystal structures of two dark states of the photoreceptor protein from the bacterium *Thermus thermophilus* were determined: CarH bound to AdoCbl and CarH bound to a large portion of the cognate DNA operator (and AdoCbl); a light state was also determined in which CarH was bound to cobalamin in which the Co-C bond had been broken. The breaking of the Co-C bond of Ado-Cbl acts as a trigger for the regulatory switch that allows the transcription of DNA. In the two dark states AdoCbl is bound to a conserved histidine from CarH, which displaces the lower 5,6-dimethylbenzimidazole ligand of AdoCbl. In the light state the 5'-deoxyadenosyl group of AdoCbl is replaced by a second histidine from CarH, giving a bis-histidine cobalamin and 4',5'-anhydroadenosine. Genes for B_{12}-dependent photoreceptors are widespread in bacteria. Control of DNA transcription may represent an evolutionarily ancient function of AdoCbl, possibly pre-dating its function as a protein cofactor.

Significance:

- A new function for adenosylcobalamin, a light-sensitive form of vitamin B_{12} with a Co-C bond, has been discovered in bacteria
- Some non-photosynthetic bacteria use adenosylcobalamin as a cofactor for the protein CarH, which controls DNA transcription
- Three X-ray crystal structures of CarH have been determined: bound to adenosylcobalamin, DNA and after light exposure
- A mechanism of action for CarH, based on its structure and on model reactions of vitamin B_{12}, is proposed

Introduction

It is appropriate that in 2015, the UNESCO year of light, the X-ray crystal structure of a novel coenzyme B_{12}-dependent photoreceptor protein was reported.[1] Vitamin B_{12} was first discovered as cyanocobalamin (CNCbl), an inactive form of the vitamin, as an artefact of the isolation procedure, which utilised cyanide. (For reviews on vitamin B_{12}, see [2–5]). The coenzyme forms of vitamin B_{12} – adenosylcobalamin (AdoCbl) and methylcobalamin (MeCbl) (Figure 1) – are extremely unstable to light and were converted to CNCbl by photolysis during the work-up procedure of their original isolation. This sensitivity to light is considered to be a great nuisance by researchers because all experiments involving coenzymes must be performed in the dark or under dim red light. However, until recently, the photosensitivity of B_{12} coenzymes appeared to have no physiological function.[1,6,7]

Humans have only two vitamin B_{12}-dependent enzymes: methylmalonyl-coenzyme A (CoA) mutase which utilises AdoCbl as its coenzyme and methionine synthase which utilises MeCbl. Methylmalonyl-CoA mutase, an example of an isomerase enzyme, converts methylmalonyl-CoA to succinyl-CoA in the oxidation of odd-chain fatty acids from the degradation of isoleucine and valine. Methionine synthase, an example of a methyl transfer enzyme, converts homocysteine to methionine. Humans (and other mammals) are unable to synthesise cobalamins and rely on dietary sources for this essential vitamin. Bacteria are much more versatile and some can synthesise cobalamins *de novo* as well as use them in a much wider variety of enzyme reactions. Bacteria utilise MeCbl in the fixation of carbon dioxide through the acetyl-CoA pathway, methanogenesis and in a variety of methylation reactions, including some involved in the synthesis of the corrinoid precursors of cobalamins. AdoCbl-requiring enzymes in bacteria include several isomerases, amino mutases, diol dehydratase, ethanolamine ammonia lyase and a ribonucleotide reductase.

The mechanism of action of AdoCbl-requiring enzymes involves homolysis of the Co-C bond of AdoCbl to give cobalamin in the Co(II) oxidation state and the 5'-deoxyadenosyl free radical. This reaction is similar to that taking place when free AdoCbl is photolysed. For AdoCbl-requiring enzymes, the 5'-deoxyadenosyl free radical takes part in rearrangement of substrate to product. For photolysis of free AdoCbl in the absence of oxygen, the 5'-deoxyadenosyl radical cyclises to give 5',8-cycloadenosine[8] and in the presence of oxygen it gives adenosine 5'-aldehyde and 5'-peroxyadenosine (Scheme 1)[9-11].

Dorothy Hodgkin determined the X-ray crystal structure of vitamin B_{12} in 1955.[12] This structure was part of the research for which she was awarded a Nobel prize in chemistry. The structures of a large number of cobalamins, including AdoCbl and MeCbl, were subsequently determined. This determination was followed, from the mid-1990s, by the protein X-ray crystal structures of first the MeCbl-binding subunit of methionine synthase[13] and then a large variety of other cobalamin-dependent enzymes, including methylmalonyl-CoA mutase[14].

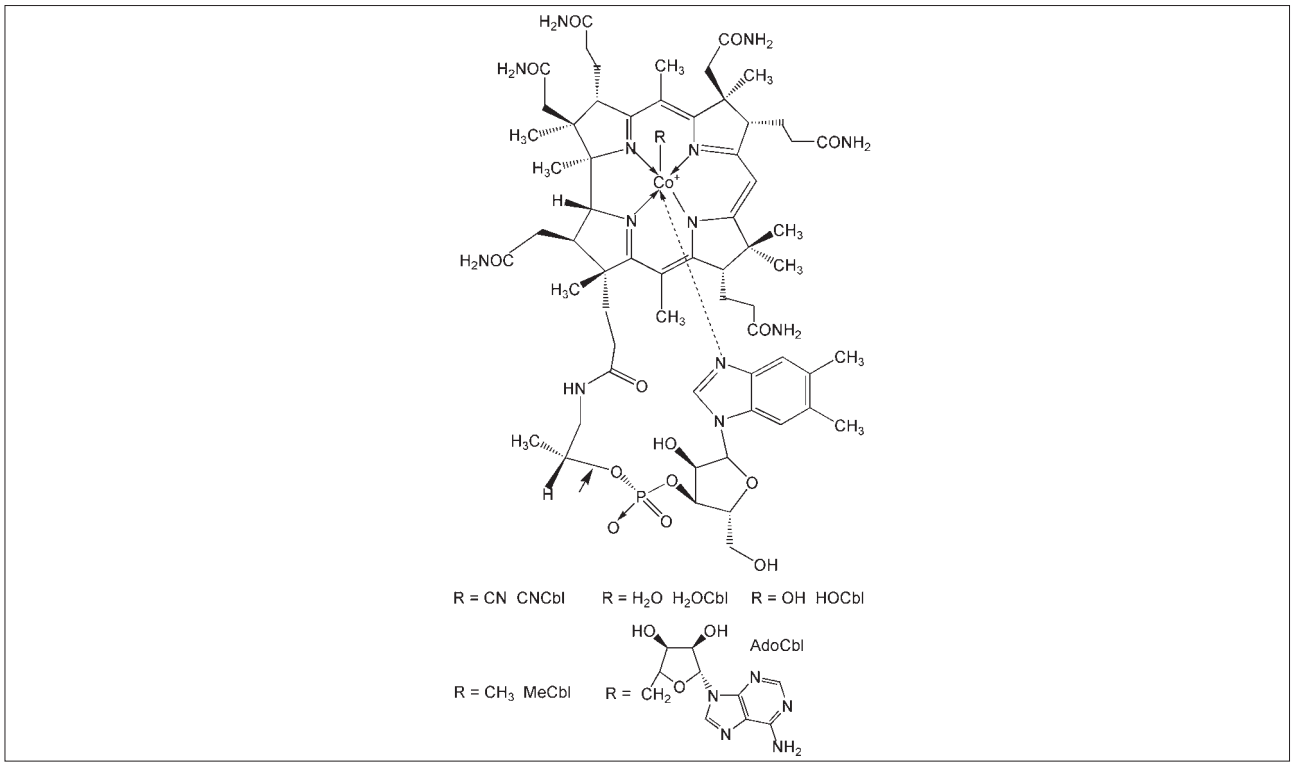
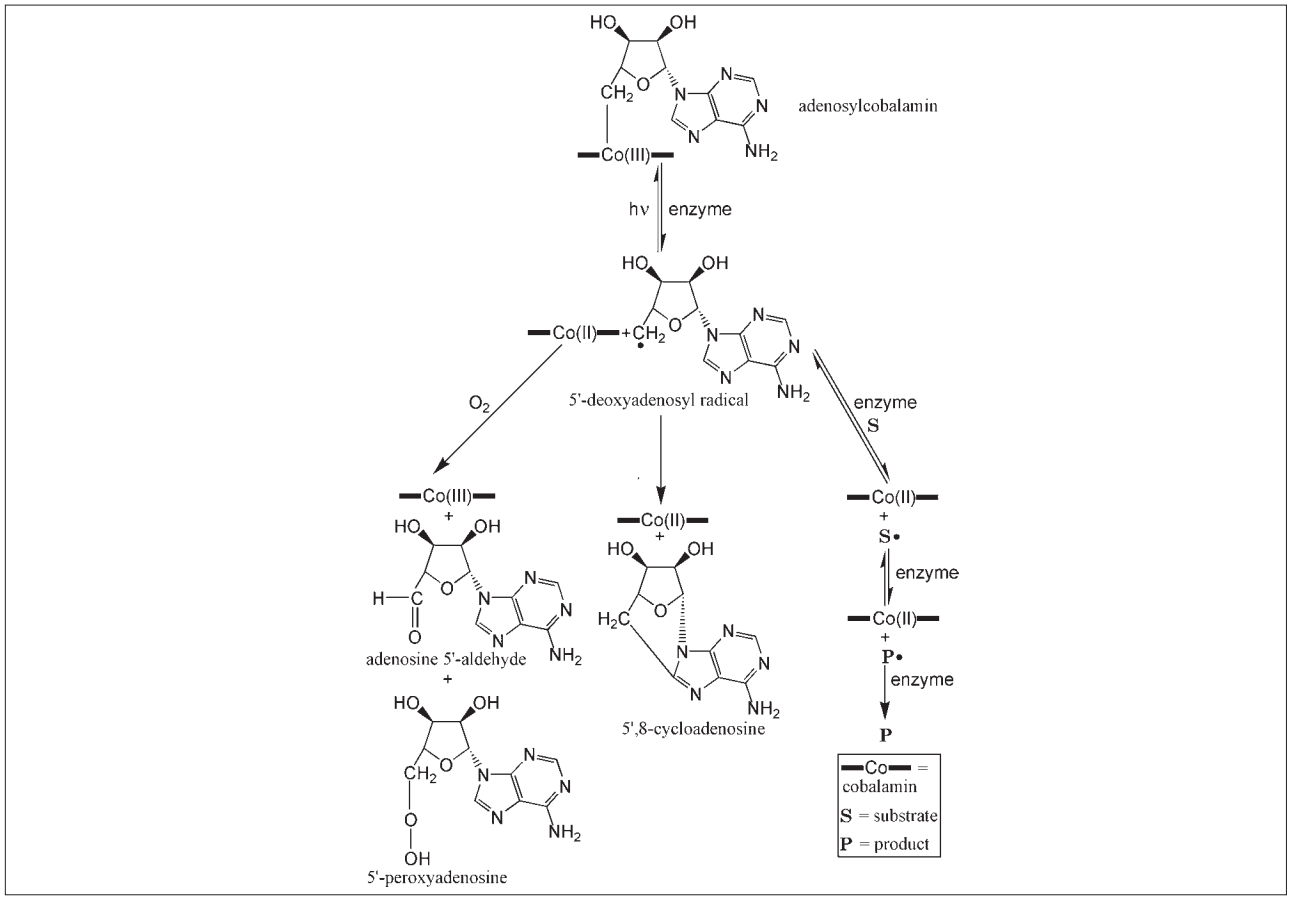


Figure 1: Structures of cobalamins. Vitamin B_{12} or cyanocobalamin (R=CN, CNCbl), aquocobalamin (R=H_2O, H_2OCbl), hydroxocobalamin (R=OH, HOCbl), methylcobalamin (R=CH_3, MeCbl) and adenosylcobalamin, also known as coenzyme B_{12} (R=5'-deoxyadenosyl, AdoCbl). Cobalamins are shown in the 'base-on' form in which the 5,6-dimethylbenzimidazole base (dbzm) is coordinated to Co, but they can also exist in the 'base-off' form in which dbzm is not coordinated to Co and may be replaced by another ligand. H_2OCbl and HOCbl are found in an acid/base-dependent equilibrium and both are present at physiological pH. Hydrolysis of the bond at → gives the corresponding cobinamide.



Scheme 1: Homolysis of the Co-C bond of adenosylcobalamin (AdoCbl).

X-ray crystallography has shown that B_{12} enzymes can bind AdoCbl in two ways: in the 'base-off' form whereby a histidine from the protein displaces the 5,6-dimethylbenzimidazole (dbzm) base, as in methionine synthase[13] and methylmalonylCoA mutase[14] and in a 'base-on' form as in diol dehydratase and ethanolamine ammonia lyase[5]. X-ray crystallography has been invaluable in providing insight into the structures and reaction mechanisms of B_{12} enzymes.

AdoCbl is also known to interact directly with messenger RNA in riboswitches. (A riboswitch is a mRNA which interacts directly and selectively with a small molecule.[15]) One of the first riboswitches to be discovered was the *E. coli btuB* mRNA,[16] which interacts selectively with AdoCbl in order to control the synthesis of the BtuB transmembrane protein, which transports corrinoids across the outer membrane of the bacteria. At low AdoCbl concentrations the riboswitch is 'on', and at high concentrations the switch is 'off'. In the 'off' position, AdoCbl binds to the mRNA, changing its three-dimensional structure, preventing the association of mRNA with the ribosome and stopping the synthesis of BtuB. AdoCbl riboswitches are widespread in bacteria and are involved in the synthesis and transport of cobalamins and the transport of metal ions.[16,17]

Adenosylcobalamin-dependent photoreceptor proteins in bacteria

Recently, a new group of photoreceptor proteins that uses AdoCbl to sense light was discovered in certain non-photosynthetic bacteria.[6,18-27] These bacteria produce carotenoids – yellow, orange or red pigments – which protect them from photo-oxidative damage by quenching the singlet oxygen free radicals produced by absorption of energy from light.[19] The photoreceptor proteins were first discovered in the bacterium *Myxococcus xanthus*,[6,18-20] which turns red in the presence of light, as a result of the synthesis of carotenoids, but is pale yellow in the dark. However, carotenoids have also been found in a variety of other bacteria, including *Streptomyces coelicolor*[21-23], *Thermus thermophilus*[24-26] and *Bacillus megaterium*[27]. Homologous sequences to the photoreceptor protein are found in the genomes of many bacteria and these proteins are probably widely distributed in non-photosynthetic bacteria.[6,17,19,22,25,26] The photoreceptor proteins were named CarH in *M. xanthus*[20], TtCarH[25] or LitR[24] in *T. thermophilus* and LitR[27] in *B. megaterium*, but, for simplicity, and because the photoreceptor proteins are all very similar, in this review I shall use the designation CarH for the protein and *carH* for the gene.

The non-photosynthetic bacteria in which CarH has been identified vary widely in their habitat and metabolism. *M. xanthus* is a rod-shaped Gram-negative predatory bacterium. It is unable to synthesise B_{12} but can obtain B_{12} from its food and convert it to AdoCbl, presumably using an ATP:corrinoid adenosyltransferase enzyme because it has a gene encoding this enzyme.[6] *T. thermophilus* is a Gram-negative, extremely thermophilic bacterium (extremophile).[28,29] *S. coelicolor* is a Gram-positive filamentous bacterium.[21,23] *B. megaterium* is a Gram-positive endospore-forming soil bacterium.[27] *T. thermophilus*[28,29], *S. coelicolor*[21,23] and *B. megaterium*[30] are all capable of synthesising cobalamins *de novo*. *M. xanthus* and *T. thermophilus* are the bacteria that have been studied most intensively. In *M. xanthus,* the *carH* gene is found on the chromosome[19] but in *T. thermophilus* it is found, together with genes for the later stages of B_{12} biosynthesis, on the megaplasmid, which is a large extrachromosomal element.[24]

The monomer of the photoreceptor protein CarH was found, by analysis of its primary structure, to consist of two regions joined by a short flexible linker chain: a C-terminal (carboxy) region, which can bind to cobalamins, and an N-terminal (amino) region, which can bind to DNA. The C-terminal end detects light and the N-terminal end allows gene expression for the production of carotenoids. The amino acid sequence at the N-terminal end of CarH is very similar to that of the DNA-binding domain of the N-termini of a family of transcriptional activators, known as the MerR-like proteins, which mediate responses to stress arising from exposure to toxic compounds or organic free radicals.[20] The amino acid sequence of the carboxyl region of CarH is very similar to that of the MeCbl-binding domain of methionine synthase and shows a typical cobalamin-binding fingerprint sequence.[6,13,18,19,31,32] However, the prosthetic group of CarH is AdoCbl, rather than MeCbl, as in methionine synthase.[6] AdoCbl is bound to CarH through displacement of dbzm by the imidazole group of a conserved histidine (His193 in *M. xanthus* and His177 in *T. thermophilus*), similarly to the binding of MeCbl in methionine synthase.[6]

The active form of the CarH receptor protein is a tetramer containing four molecules of AdoCbl (Figure 2). CarH monomers oligomerise to form the tetramer only when AdoCbl is present. In the dark, the CarH tetramer binds to the DNA operator controlling carotenogenesis and prevents the action of RNA polymerase, thus blocking transcription of the proteins needed for synthesis of carotenes. In the presence of light at wavelengths at which AdoCbl absorbs light (360 nm, 438 nm or 540 nm), AdoCbl is photolysed, the tetramer dissociates into monomers and the gene repression is released. This process occurs only if the conserved histidine from the cobalamin-binding fingerprint sequence is present; mutants lacking this histidine, when histidine is replaced by alanine, do not show AdoCbl-dependent oligomerisation.[6]

It was later discovered that vitamin B_{12} also has a gene regulation function in the photosynthetic purple bacteria *Rhodobacter capsulatus*.[33] In the presence of light but the absence of oxygen, *R. capsulatus* produces large amounts of bacteriochlorophyll, and photosynthesis takes place through an anaerobic pathway. However, *R. capsulatus* can also grow aerobically in the dark.[34] In the presence of sufficient oxygen, when the anaerobic pathway is not needed, the pathway is repressed because it also produces destructive singlet oxygen.[17,33] The protein CrtJ represses the synthesis of bacteriochlorophyll and is linked to the protein AerR, which is also an aerobic repressor of photosynthesis genes.[35] The AerR protein uses AdoCbl or MeCbl as its cofactor but does not bind to either.[17,33] It was proposed (Figure 3) that AdoCbl or MeCbl is converted to hydroxocobalamin (HOCbl) in the presence of light, which is followed by binding of HOCbl to AerR, then binding of HOCbl/AerR to the gene repressor CrtJ, and finally release of the repression of the genes for synthesis of bacteriochlorophyll so that anaerobic photosynthesis can take place. AerR does not bind to AdoCbl or MeCbl but does bind to their photolysis product, HOCbl, and can therefore also be considered as a light-sensing protein.[17,33]

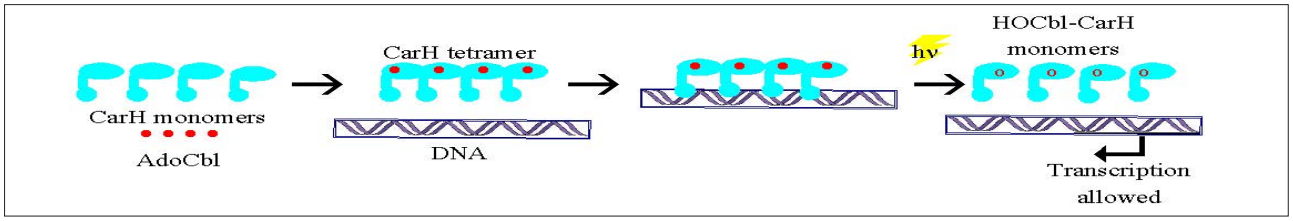


Figure 2: Model for the response to light of the AdoCbl regulatory switch in *Myxococcus xanthus*.[6] In the dark, AdoCbl binds to CarH monomers to give the CarH tetramer and the CarH tetramer binds to DNA, preventing transcription. In the presence of light, AdoCbl is photolysed and CarH dissociates, allowing transcription.

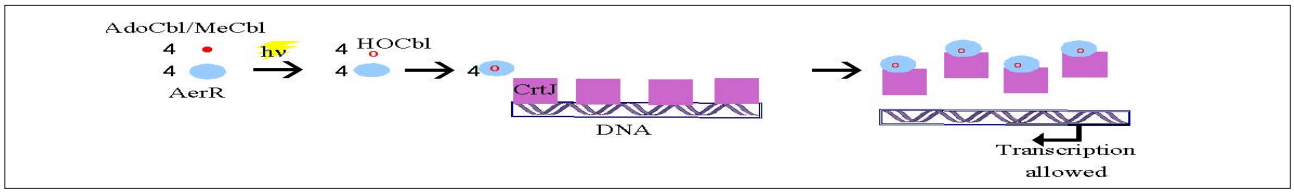


Figure 3: Model for the response to light of the B_{12}-dependent regulatory switch in *Rhodobacter capsulatus*.[33] In the dark, the CrtJ repressor binds to DNA, preventing transcription. In the presence of light, AdoCbl or MeCbl is photolysed to give HOCbl which binds to AerR. The AerR–cobalamin complex binds to CrtJ and the CrtJ dissociates, allowing transcription.

The AerR protein is similar to CarH in that it is homologous to CarH and methionine synthase, has a typical cobalamin binding fingerprint sequence and contains a histidine (His145) which binds to the lower position of B_{12}, displacing dbzm. It is different in that it binds only HOCbl (rather than AdoCbl or MeCbl) and does not contain any DNA-binding region, but instead interacts with the protein CrtJ. AerR binds to HOCbl very strongly, much more strongly than expected from an ionic interaction between cobalamin and a single histidine on AerR. It was proposed that the upper position of the cobalamin is occupied by a second histidine (His10) from AerR and that this bond is stronger than that to the lower histidine.[33]

X-ray crystal structure of the CarH photoreceptor protein

In 2015, Jost, Drennan and coworkers[1] determined the X-ray crystal structure of the photoreceptor CarH from *T. Thermophilus* in three different states: two dark states, in one of which CarH is free and in the other is attached to DNA and a state in which CarH has dissociated from DNA after exposure to visible light (Figure 4).

In the free dark state, CarH is a tetramer with one molecule of AdoCbl bound to each monomer.[1] Each monomer has one domain that attaches to DNA at the N-terminal end and a second domain that binds the light-sensing AdoCbl molecule at the C-terminal end. The N-terminal domain consists of a DNA-recognition helix and a β-hairpin wing. The C-terminal domain consists of a four-helix bundle and a Rossman fold cobalamin-binding region. AdoCbl, in which the bond length of the Co-C bond is 2.0 Å, similarly to that of free AdoCbl (2.030 Å)[36], is located between the four-helix bundle and the Rossman fold[1]. The upper ligand, 5'-deoxyadenosyl, is orientated towards the four-helix bundle, close to a tryptophan residue, Trp131, and the lower ligand dbzm has been displaced by a histidine ligand (His177) from the Rossman fold. The C-terminal domain is rigid but the N-terminal domain is more flexible. The tetramer assembles itself from the monomers only if AdoCbl is present. The tetramer can be considered as a dimer of dimers, in which each dimer has its monomers in a head-to-tail orientation with respect to the AdoCbl-binding domain. The monomers within the dimer are held together by many hydrogen bonds and electrostatic interactions from a variety of side-chains and from the 5'-deoxyadenosyl ligand. The dimers are joined through their AdoCbl-binding domains in a staggered fashion to give the tetramer, involving Gly160 and Gly192 on each dimer. The C-terminal domains of the monomers form the centre of the tetramer with the N-terminal domains on the exterior.[1]

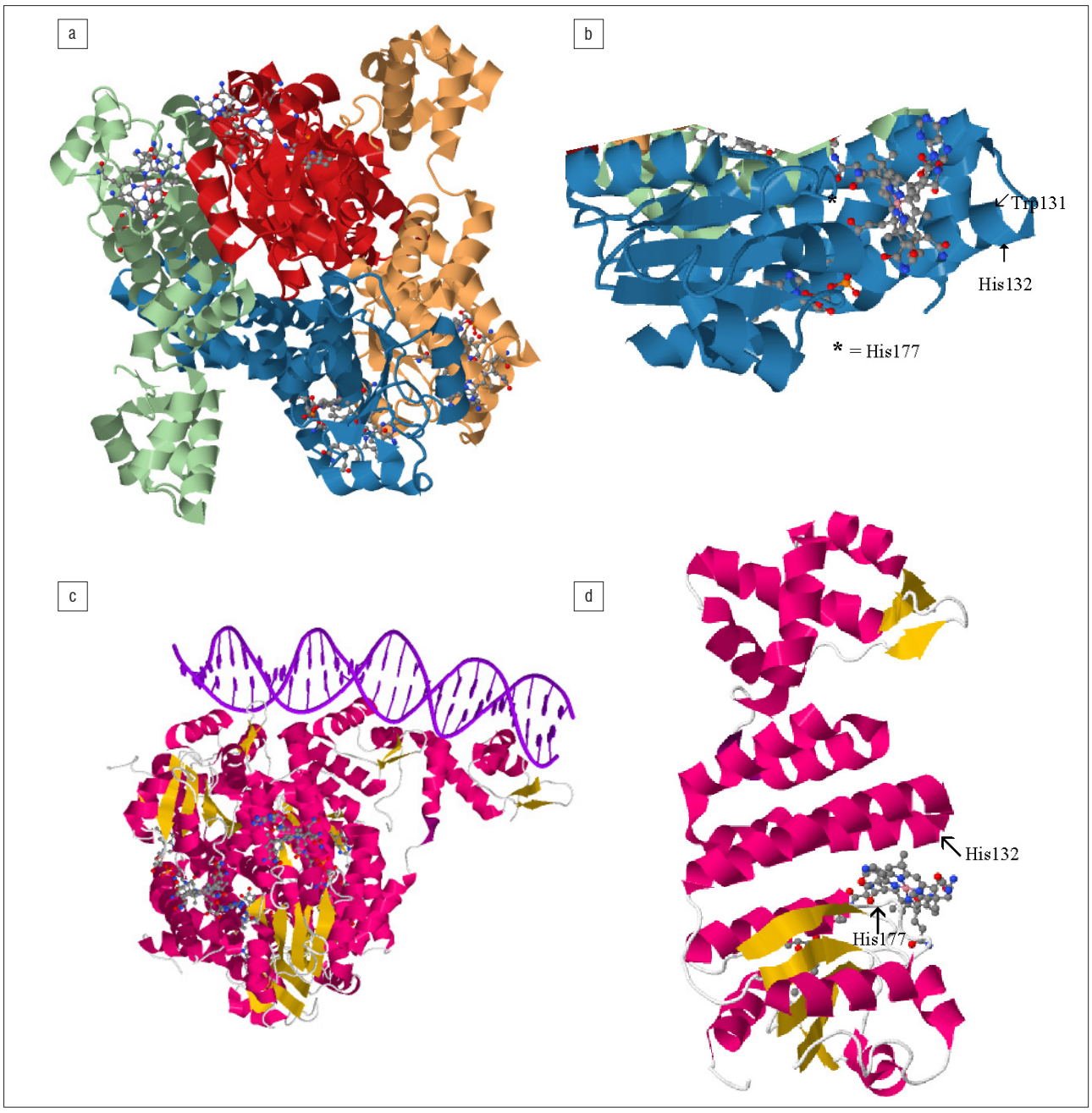
It is interesting that the primary structure of the AdoCbl-binding domain of CarH is similar to that of the MeCbl-binding domain of methionine synthase, rather than to any of the enzymes using AdoCbl as a prosthetic group, especially as the adenosyl group is much larger than the methyl group.[1,6,13,18,19,31,32] However, the cobalamin binding pocket in CarH is bigger than that in methionine synthase because the four-helix bundle is situated 2.5 Å further away from the cobalamin and the leucine in methionine synthase is replaced by a smaller valine (Val138) residue. The pocket also provides more opportunities for hydrogen bonding by replacing a valine with glutamic acid (Glu141) and for polar interactions by replacing a valine with histidine (His142) and phenylalanine with tryptophan (Trp131).[1]

In the model for the AdoCbl-based regulatory switch (Figure 2), the CarH tetramer binds to its cognate DNA operator, a 30-bp (base pair) portion of the region of DNA between the gene encoding CarH and the gene encoding CrtB (the operator controlling carotenogenesis). In the crystal structure of the second dark state, the CarH tetramer is bound to a large piece (26 bp) of the cognate DNA operator.[1] Three of the four DNA-binding domains of the CarH tetramer bind to DNA, with the fourth domain being disordered and not visible in the crystal structure. The three detectable DNA-binding domains all face in the same direction but otherwise the structure is similar to that of the free dark state. All three visible DNA-binding domains are important in binding CarH to DNA, forming a variety of hydrogen bonds and electrostatic interactions between the protein and the phosphate backbone of the DNA. In addition, the recognition helix of each DNA domain on CarH inserts into the DNA major groove and a histidine from the β-hairpin wing fits into the DNA minor groove. The recognition helix of the middle DNA domain covers the promoter -35 element for the gene corresponding to the σ^A-associated bacterial RNA polymerase, in which σ^A is the protein subunit necessary for initiation of RNA synthesis[37], and thus blocks access of the RNA polymerase to the promoter. (The promoter -35 element is so called because it is 35 nucleotides upstream, or counting backwards, from the transcription start site of σ^A.[38])

In contrast to the dark-state tetramer, the light-state CarH protein is monomeric and contains cobalamin, but has lost the 5'-deoxyadenosyl group.[1] When comparing the light-state CarH monomer with the corresponding dark-state monomer, it can be seen that the Co-C bond of AdoCbl has been broken and the 5'-deoxyadenosyl group is absent. The helical bundle, including Trp131, has moved into the vacant space left by the 5'-deoxyadenosyl group and is positioned very much closer to cobalamin. Very importantly, a histidine, His132, on the helical bundle has shifted into a suitable position to coordinate with Co(III) on the cobalamin through its imidazole side-chain.

Mechanism of action of the CarH regulatory switch

The X-ray crystal structures of the three CarH proteins are consistent with the AdoCbl-based regulatory switch proposed in Figure 2, but give a much more detailed insight into its mechanism. In the dark state, the lower ligand (dbzm) of AdoCbl is displaced by the imidazole group of His177 and is thus bound through this histidine to CarH. In the His177 $\rightarrow$ Ala mutant, this step is blocked and the regulatory switch cannot function.[1,6] In the dark, when carotenogenesis is not needed, the CarH tetramer binds to the DNA operator, preventing the transcription of carotenogenesis proteins. The X-ray crystal structures show that the DNA-binding domains of CarH are conveniently located on the outside of the tetramer, whereas the AdoCbl-binding domains are each buried in a deep pocket. In the presence of visible light, when carotenogenesis is of advantage to the bacteria, the Co-C bond of the bound AdoCbl is broken and the 5'-deoxyadenosyl moiety drifts away. The change in conformation on the breaking of the Co-C bond causes the movement of the helical bundle into the pocket, bringing His132 into position to bind to Co(III) through its imidazole side-chain. The change of conformation disrupts the head-to-tail dimer interaction and disassembles the tetramer, causing CarH to dissociate from the DNA promoter and relieving the blockage of access of RNA polymerase to the promoter so that transcription is allowed.[1]



Source: Images from the RCSB PDB (www.rcsb.org)*; PDB IDs: 5C8D, 5C8A, 5C8E and 5C8F.*[1]

Figure 4: Structures of CarH. (a) CarH dark-state tetramer, showing the four identical subunits in four colours, and four molecules of AdoCbl (Co, coral; C, grey; N, blue; O, red; P, orange), one bound to each subunit (PDB ID: 5C8D). (b) CarH dark-state monomer, showing one of the four identical subunits (blue) and one molecule of AdoCbl (colours as in (a)). The truncated form of CarH (PDB ID: 5C8A), which consists of only the light-sensitive C terminal domain, is depicted in order to show the binding of AdoCbl more clearly. Trp131 and His132 are indicated by arrows and His177 by an asterisk. (c) CarH dark-state tetramer bound to 26 base pairs of the cognate DNA operator (DNA, purple; α helix, pink; β sheet, yellow; all other structures, white; colours of the AdoCbl molecules are as in (a)). The three DNA-binding domains, all facing in the same direction, can be seen at the bottom of the figure (PDB ID: 5C8E). (d) CarH light-state monomer (colours as in (c)). His132 and His177 are indicated by arrows and Trp 131 is behind His132 (PDB ID: 5C8F)

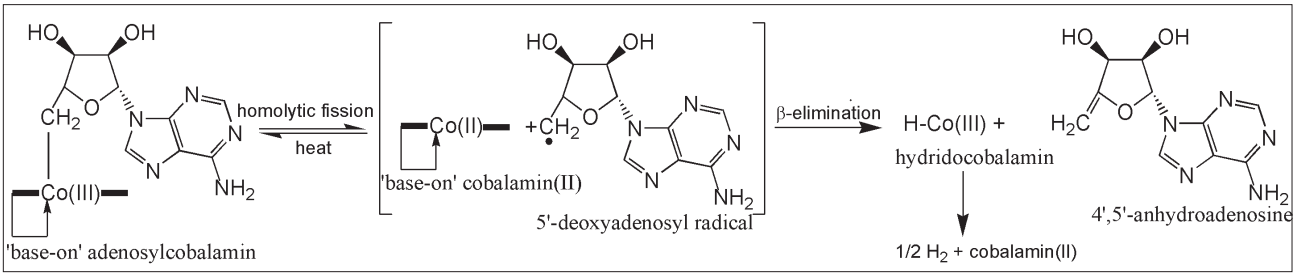
In conjunction with X-ray crystallography, Jost et al.[1,7] have used other techniques to answer questions about the mechanism of photolysis of CarH-bound AdoCbl. Two of these questions are:

1. What are the products of photolysis of CarH-bound AdoCbl and how exactly is the Co-C bond in CarH-bound AdoCbl broken?
2. How is the Co(III) in the light-state CarH protein bonded to the second histidine (His 132) on the protein?

Both questions have already generated controversy and will be considered further here.

CarH-bound AdoCbl: Products of photolysis and the Co-C bond

The UV-visible spectrum of the CarH free dark state is very similar to that of free AdoCbl and is completely consistent with the imidazole group of histidine being the lower ligand.[1]



Scheme 2: Homolytic fission followed by β-elimination for the thermolysis of adenosylcobalamin in glycerol.

After photolysis in the presence of oxygen the CarH-bound cobalamin is in the Co(III) oxidation state[1,7] but in the absence of oxygen the cobalamin is in the Co(II) oxidation state, as shown by UV-visible spectroscopy and electron spin resonance spectroscopy.[7] Exposure of the CarH-bound cobalamin in the Co(II) oxidation state to oxygen, after anaerobic photolysis, gives the same Co(III) product as does aerobic photolysis. However, the organic product of photolysis of CarH-bound AdoCbl was completely unexpected. Jost, Drennan and coworkers[7] have shown unequivocally by liquid chromatography–mass spectrometry and nuclear magnetic resonance spectroscopy that the organic photolysis product of AdoCbl-bound CarH under both anaerobic and aerobic conditions is solely 4',5'-anhydroadenosine. The product 4',5'-anhydroadenosine had never been observed previously in the photolysis of AdoCbl (see above). However 4',5'-anhydroadenosine has been observed in the thermolysis of AdoCbl in glycerol[39-41], which could provide a model for the photolysis of CarH-bound AdoCbl.

Garr and Finke[39,41] found that, in the highly viscous solvent glycerol, thermolysis of AdoCbl at 110 °C gives rise to 5% 4',5'-anhydroadenosine (in addition to 5',8-cycloadenosine and 5'-deoxyadenosine), but in the less viscous solvent, ethylene glycol, no 4',5'-anhydroadenosine is seen. Thermolysis of adenosylcobinamide (the 5-coordinate analogue of AdoCbl in which the nucleotide base has been removed) in ethylene glycol gives 4',5'-anhydroadenosine as a major product (33%).[40,41] For both AdoCbl and adenosylcobinamide, the corrinoid partner of 4',5'-anhydroadenosine is Co(II).[39-41] No 4',5'-anhydroadenosine is seen in thermolysis of AdoCbl[42] or adenosylcobinamide[43] in aqueous solution. A viscous solvent, which can act as a strong cage for Co(II) and the 5'-deoxadenosyl radical, appears to be necessary for the comparatively slow formation of 4',5'-anhydroadenosine. Garr and Finke[39-41] proposed that, while in close proximity in the cage, the Co(II) and the 5'-deoxyadenosyl radical (generated by homolytic fission) undergo a β-elimination reaction to give hydridocobalamin and 4',5'-anhydroadenosine. Hydridocobalamin then rapidly decomposes to Co(II) and hydrogen (Scheme 2).[39-41,44]

Based on the results of Garr and Finke[39-41], Jost and Drennan and coworkers[7] have proposed the following CarH photolysis mechanisms (Scheme 3, Paths 1 and 2). In Path 1 of Scheme 3, CarH-bound AdoCbl first undergoes photolysis to give Co(II) and the 5'-deoxyadenosyl radical by homolytic fission of the Co-C bond. Then, while in close proximity in the extra strong cage provided by the protein, β-elimination between Co(II) and the 5'-deoxyadenosyl radical takes place to give the Co(III) hydride (still bound to the protein) and 4',5'-anhydroadenosine. Clearly, the CarH protein would provide a much stronger cage than a viscous solvent, which would account for 4',5'-anhydroadenosine being the sole product of photolysis of CarH-bound AdoCbl.[7] There is a large amount of evidence in favour of Path 1 and this path is favoured by Jost and Drennan and colleagues[7]. Cage effects by proteins have been previously observed in the AdoCbl-dependent glutamate mutase[45,46] and in the MeCbl-binding domain of methionine synthase[47]. In addition to thermolysis of AdoCbl and adenosylcobinamide[39-41], homolytic fission followed by radical-mediated β-elimination is well documented in alkylcobalamins and alkylcobinamides[48-50]. The generation of 4',5'-anhydroadenosine (rather than more reactive species as in photolysis in vitro) as the organic partner of CarH photolysis may represent a safety mechanism in ensuring that the reactive 5'-deoxyadenosyl radical is not released.[1,7]

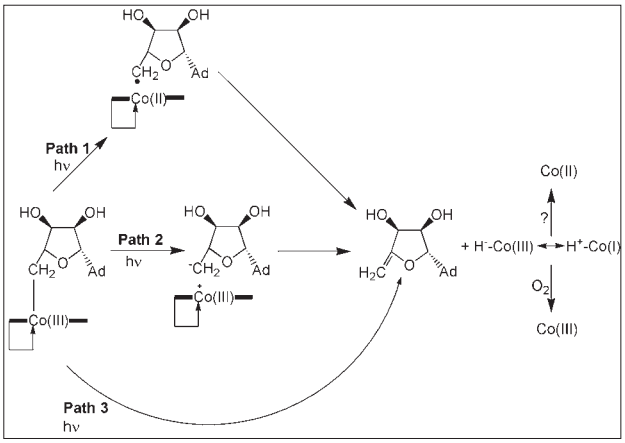
The intermediate hydridocobalamin, formulated

$$Co(I)\text{-}H^{+} \leftrightarrow Co(II)\text{-}H \leftrightarrow Co(III)\text{-}H^{-},$$

is an unstable species that rapidly disproportionates to give Co(II) and molecular hydrogen under anaerobic conditions[44] and is presumably oxidised to Co(III) under aerobic conditions. Hydridocobalamin is difficult to characterise because it is extremely unstable. However, from cyclic voltammetry[51], hydridocobalamin has $pK_a \approx 1$ and is protonated on the 5,6-dimethybenzimidazole base as well as on the Co[51-53]. Thus, hydridocobalamin can be considered as protonated Co(I) and can act as a Brönsted acid to protonate water:

$$Co(I)\text{-}H^{+} + H_2O \rightarrow Co(I) + H_3O^{+},$$

or possibly an amino acid side-chain on the protein.



Ad, adenine base

Scheme 3: Paths proposed for CarH photolysis.

Some doubt has been expressed about the structure of hydridocobalamin because 'hydridocobaloxime', which was proposed to have the same structure as hydridocobalamin[54], has been found to be a dimeric Co(II) compound with a long Co-Co bond[55]. Cobaloximes are not good models for cobalamins in this case because they are relatively flat and can dimerise more readily than cobalamins. Cobalamins are prevented from dimerising in the same way by steric hindrance from their side-chains preventing a close enough approach of the two Co atoms.

Path 2 in the scheme of Jost et al.[7] involves initial heterolytic cleavage of the Co-C bond to give Co(III) and the 5'-deoxyadenosyl anion, followed by β-elimination to give 4',5'-anhydroadenosine and hydridocobalamin. Kutta et al.[56] propose a variant of Path 2 in which heterolysis of the Co-C bond gives the 5'-deoxyadenosyl anion and either hydridocobalamin or a five-coordinate positively charged Co(III) as intermediates on the route to 4',5'-anhydroadenosine and cob(II)alamin. It would seem that Path 2 would be highly unlikely. The 5'-deoxyadenosyl anion species appears to be completely unprecedented; a search of SciFinder produced no references compared with 284 for the 5'-deoxyadenosyl radical.[57] An

anion such as the 5'-deoxyadenosyl anion would be expected to be extremely unstable and would immediately rearrange or decompose to give a more stable species. This expectation has been observed in at least one case. A model reaction for heterolytic cleavage of the Co-C bond in CarH-bound AdoCbl is provided by the thermolysis of AdoCbl in aqueous solution at pH 7.0 and 85 °C. This reaction proceeded by 10% heterolysis, in which the organic products of heterolysis were only adenine and 2,3-dihydroxypentenal, which are decomposition products of 5'-deoxyadenosyl.[42] The presence of exclusively 4',5'-anhydroadenosine (and no adenine) as the organic product of CarH photolysis strongly suggests that Path 2 is not operative in CarH photolysis.

Yet another path (Path 3) – concerted β-elimination by migration of a hydride ion (H^-) to give hydridocobalamin and 4',5'-anhydroadenosine directly (that is a reaction not proceeding through a radical pair intermediate) – is theoretically possible. This type of reaction seems unlikely because it generally requires a vacant coordination site on the metal[58] and has no precedent in organocobalamins.

How is Co(III) in the light-state CarH protein bonded to His 132?

After photolysis has taken place, the cobalamin moiety binds very strongly to light-state CarH, with the wild-type CarH forming a very stable complex with the cobalamin.[1] The cobalamin-dependent light-sensing protein AerR in *Rhodobacter capsulatus* similarly forms a very stable complex with cobalamin and it has been proposed that this complex contains cobalamin bound to two histidines from the protein at both the top and bottom positions.[17,33] However, this type of coordination of histidine is controversial because it has not been observed for free cobalamin (Marques et al.[59] and personal observations). It also means that it is not possible to directly compare the UV-visible spectrum of the light-state CarH protein with that of free bis-histidine cobalamin in aqueous solution. The equilibrium constant for coordination to Co(III) of the first histidine ($\log K_1$ 4.30) is favourable but coordination of the second histidine is much more difficult ($\log K_2 < -1$),[59] so that the spectrum of the bis-histidine complex is not observed.[59] However, the UV-visible spectra of cobalamin with histidine or imidazole in the upper position and dbzm in the lower position are almost indistinguishable,[59] and the UV-visible spectrum of (mainly) the bis-imidazole complex of cobalamin can be observed as $\log K_1 = 4.59$ and $\log K_2 = 0.6$.[59] The UV-visible spectrum of the light-state CarH protein is very similar to the UV-visible spectrum of the Co(III) cobalamin complex with two imidazole ligands binding through their N atoms in the upper and lower positions[1,59] and is consistent with the substitution of cobalamin by two histidines from the protein[1]. Presumably, the cobalamin is constrained by the protein so that substitution of a second histidine becomes possible. The extreme stability of light-state CarH[1] and AerR bound to cobalamin[33], both of which show a cobalamin adduct in mass spectrometry[1,33], suggests that the top histidine has a covalent bond to Co(III), rather than the ionic bond in the lower position[1,33]. It is possible that the top histidine is bound in the anion form, because imidazole bound as the anion has a larger logK than that for the neutral form (logK 4.60 rather than logK 4.30),[59] which might account for the great stability of the light-state CarH protein. If His132 in the light-state CarH protein is mutated to alanine so that only one histidine can bind, the UV-visible spectrum of the corresponding light-state CarH protein is very similar to the UV-visible spectrum of Co(III) cobalamin with only one imidazole[1,59] or histidine[59] ligand and it is less stable than the wild-type light CarH protein[1].

Photochemistry of CarH

In an attempt to determine a more detailed photochemical mechanism for CarH, Kutta et al.[56] performed photoexcitation experiments on the AdoCbl-bound CarH tetramer (ground-state CarH) and used transient UV-visible absorption spectroscopy on a femtosecond (10^{-15} s) to second timescale to follow the intermediates on the way to the light-state CarH product. They observed, in addition to Co(II) (and the presumed 5'-deoxyadenosyl radical), transient absorption spectra of at least eight intermediates and assigned these to possible structures, based on a model of the reaction pathway from ground-state CarH to light-state CarH. In their model, Intermediate A is the initial photoexcited state which, in their major pathway, immediately decays back to ground-state CarH either directly or through the Co(II)/5'-deoxyadenosyl radical pair, B. In their major pathway, the formation of the Co(II)/5'-deoxyadenosyl radical pair is unproductive, leading only to recombination to give AdoCbl. In their minor pathway, A converts to C, which has a spectrum that can be interpreted as a metal-to-ligand charge transfer complex, consisting of Co(III) with a partial positive charge and an incipient 5'-deoxyadenosyl anion. C then converts to D* (in the text) or C* (in their Figure 6),[56] which is interpreted as a five-coordinate, positively charged Co(III) together with a 5'-deoxyadenosyl anion, then to D, in which the 5'-deoxyadenosyl anion has moved away from Co and has been replaced by His132, and lastly to E, which is similar to light-state CarH.[56] Alternatively, they propose that Intermediate C corresponds to hydridocobalamin.[56] The UV-visible spectrum of C is very similar in shape to that proposed for base-off hydridocobalamin,[52] but is red-shifted relative to this spectrum. It is possible that C is the base-on form of hydridocobalamin but it is dangerous to speculate in this way (see below).

Further investigation is needed to elucidate the photochemical mechanism of CarH. Firstly, transient intermediate UV-visible spectra (300–700 nm) mainly tell us about the cobalamin partner and say very little about the 5'-deoxyadenosyl partner, which absorbs below 300 nm. It would perhaps be useful to extend the investigation into the UV region. Secondly, the identification of transient cobalamin intermediates in a mechanistic pathway, based only on their UV-visible spectra, is very difficult, unless suitable spectra of known species are available for comparison. For example, it is stated[56] that the spectrum of Intermediate C appears to be similar to the S_1 state seen in the photolysis of free MeCbl[601,61]. The S_1 state in MeCbl photolysis has now been characterised as a $d/\pi \rightarrow \pi^*$ metal-to-ligand charge-transfer state[61] (consistent with theoretical density functional theory calculations)[62], but was originally considered as 'a cob(III)alamin with a very weak axial ligand'[60]. The corresponding theoretical calculations for AdoCbl have not been completed to date but steady progress is being made with such calculations.[61]

Evolutionary implications

The discovery of the AdoCbl-dependent CarH photoreceptor and of the cobalamin-dependent AerR photoreceptor has many interesting implications. Genes for photoreceptors similar to CarH are found in a large number (>120) and variety of species of bacteria,[6] including bacteria that can biosynthesise cobalamins and those that acquire cobalamins from their food. Genes similar to that for AerR are found in several species of purple photosynthetic bacteria.[33] The presence of these genes suggests that B_{12}-dependent photoreceptors are widespread in both non-photosynthetic and photosynthetic bacteria. Together with adenosycobalamin-dependent riboswitches, CarH and AerR (in combination with its partner CrtJ) may represent an evolutionarily very old function for cobalamin: interaction with and control of nucleic acids. AdoCbl may have been co-opted by proteins only at a later date. AdoCbl, which is essential for B_{12} enzymes in humans (as well as other animals and bacteria), may have first evolved as a regulatory switch for DNA and RNA.

Acknowledgements

I thank Dr Peter R. Johnson for his careful reading of the manuscript and the School of Chemistry, University of the Witwatersrand for a position as Emeritus Researcher.

References

1. Jost M, Fernández-Zapata J, Polanco MC, Ortiz-Guerrero JM, Chen PY-T, Kang G, et al. Structural basis for gene regulation by a B_{12}-dependent photoreceptor. Nature. 2015;526:536–541. http://dx.doi.org/10.1038/nature14950

2. Banerjee R, Ragsdale SW. The many faces of vitamin B_{12}: Catalysis by cobalamin-dependent enzymes. Annu Rev Biochem. 2003;72:209–247. http://dx.doi.org/10.1146/annurev.biochem.72.121801.161828

3. Kräutler B. Organometallic chemistry of B_{12} coenzymes. In: Sigel A, Sigel H, Sigel RKO, editors. Metal-carbon bonds in enzymes and cofactors. Metal ions in life sciences. Volume 6. Cambridge: Royal Society of Chemistry; 2009. p. 1–51. http://dx.doi.org/10.1039/9781847559333-00001

4. Matthews RG. Cobalamin- and corrinoid-dependent enzymes. In Sigel A, Sigel H, Sigel RKO, editors. Metal-carbon bonds in enzymes and cofactors. Metal ions in life sciences. Volume 6. Cambridge: Royal Society of Chemistry; 2009. p. 53–114. http://dx.doi.org/10.1515/9783110436587-006

5. Gruber K, Puffer B, Kräutler B. Vitamin B_{12}-derivatives – Enzyme cofactors and ligands of proteins and nucleic acids. Chem Soc Rev. 2011;40:4346–4363. http://dx.doi.org/10.1002/chin.201145267

6. Ortiz-Guerrero JM, Polanco MC, Murillo FJ, Padmanabhan S, Elías-Arnanz M. Light-dependent gene regulation by a coenzyme B_{12}-based photoreceptor. Proc Natl Acad Sci USA. 2011;108:7565–7570. http://dx.doi.org/10.1073/pnas.1018972108

7. Jost M, Simpson JH, Drennan CL. The transcription factor CarH safeguards use of adenosylcobalamin as a light sensor by altering the photolysis products. Biochemistry. 2015;54:3231–3234. http://dx.doi.org/10.1021/acs.biochem.5b00416

8. Law PY, Wood JM. The photolysis of 5'-deoxyadenosylcobalamin under anaerobic conditions. Biochim Biophys Acta. 1973;331:451–454. http://dx.doi.org/10.1016/0005-2787(73)90032-4

9. Hogenkamp HPC, Ladd JN, Barker HA. The identification of a nucleoside derived from coenzyme B_{12}. J Biol Chem. 1962;237:1950–1952.

10. Hogenkamp HPC. A cyclic nucleoside derived from coenzyme B_{12}. J Biol Chem. 1963;238:477–480.

11. Schwartz PA, Frey PA. 5'-Peroxyadenosine and 5'-peroxyadenosylcobalamin as intermediates in the aerobic photolysis of adenosylcobalamin. Biochemistry. 2007;46:7284–7292. http://dx.doi.org/10.1021/bi700077v

12. Hodgkin DC, Kamper J, Mackay M, Pickworth J, Trueblood KN, White JG. Structure of vitamin B_{12}. Nature. 1956;178:64–66. http://dx.doi.org/10.1038/178064a0

13. Drennan CL, Huang S, Drummond JT, Matthews RG, Ludwig ML. How a protein binds B_{12}: A 3.0 Å X-ray structure of B_{12}-binding domains of methionine synthase. Science. 1994;266:1669–1674. http://dx.doi.org/10.1126/science.7992050

14. Mancia F, Keep NH, Nakagawa A, Leadlay PF, McSweeney S, Rasmussen B, et al. How coenzyme B_{12} radicals are generated: The crystal structure of methylmalonyl-coenzyme A mutase at 2 Å resolution. Structure. 1996;4:339–350. http://dx.doi.org/10.1016/s0969-2126(96)00037-8

15. Nahvi A, Barrick JE, Breaker RR. Coenzyme B_{12} riboswitches are widespread genetic control elements in prokaryotes. Nucleic Acids Res. 2004;32:143–150. http://dx.doi.org/10.1093/nar/gkh167

16. Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR. Genetic control by a metabolite binding mRNA. Chem Biol. 2002;9:1043–1049. http://dx.doi.org/10.1016/s1074-5521(02)00224-7

17. Klug G. Beyond catalysis: Vitamin B_{12} as a cofactor in gene regulation. Mol Microbiol. 2014;91:635–640. http://dx.doi.org/10.1111/mmi.12490

18. Cervantes M, Murillo FJ. Role for vitamin B_{12} in light induction of gene expression in the bacterium *Myxococcus xanthus.* J Bacteriol. 2002;184:2215–2224. http://dx.doi.org/10.1128/jb.184.8.2215-2224.2002

19. Elías-Arnanz M, Fontes M, Padmanabhan S. Carotenogenesis in *Myxococcus xanthus*: A complex regulatory network. In: Whitworth DE, editor. Myxobacteria: Multicellularity and differentiation. Washington, DC: ASM Press; 2008. p. 211–225. http://dx.doi.org/10.1128/9781555815677.ch12

20. Pérez-Marín MC, Padmanabhan S, Polanco MC, Murillo FJ, Elías-Arnanz M. Vitamin B_{12} partners the CarH repressor to downregulate a photoinducible promoter in *Myxococcus xanthus*. Mol Microbiol. 2008;67:804–819. http://dx.doi.org/10.1111/j.1365-2958.2007.06086.x

21. Takano H, Obitsu S, Beppu T, Ueda K. Light-induced carotenogenesis in *Streptomyces coelicolor* A3(2): Identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster. J Bacteriol. 2005;187:1825–1832. http://dx.doi.org/10.1128/jb.187.5.1825-1832.2005

22. Takano H, Beppu T, Ueda K. The CarA/LitR-family transcriptional regulator: Its possible role as a photosensor and wide distribution in non-phototrophic bacteria. Biosci Biotechnol Biochem. 2006;70:2320–2324. http://dx.doi.org/10.1271/bbb.60230

23. Takano H, Hagiwara K, Ueda K. Fundamental role of cobalamin biosynthesis in the developmental growth of *Streptomyces coelicolor* A3 (2). Appl Microbiol Biotech. 2015;99:2329–2337. http://dx.doi.org/10.1007/s00253-014-6325-z

24. Takano H, Kondo M, Usui N, Usui T, Ohzeki H, Yamazaki R, et al. Involvement of CarA/LitR and CRP/FNR family transcriptional regulators in light-induced carotenoid production in *Thermus thermophilus*. J Bacteriol. 2011;193:2451–2459. http://dx.doi.org/10.1128/jb.01125-10

25. Díez AI, Ortiz-Guerrero JM, Ortega A, Elías-Arnanz M, Padmanabhan S, De la Torre JG. Analytical ultracentrifugation studies of oligomerization and DNA-binding of TtCarH, a *Thermus thermophilus* coenzyme B_{12}-based photosensory regulator. Eur Biophys J. 2013;42:463–476. http://dx.doi.org/10.1007/s00249-013-0897-x

26. Takano H, Agari Y, Hagiwara K, Watanabe R, Yamazaki R, Beppu T, et al. LdrP, a cAMP receptor protein/FNR family transcriptional regulator, serves as a positive regulator for the light-inducible gene cluster in the megaplasmid of *Thermus thermophilus*. Microbiology. 2014;160:2650–2660. http://dx.doi.org/10.1099/mic.0.082263-0

27. Takano H, Mise K, Hagiwara K, Hirata N, Watanabe S, Toriyabe M, et al. Role and function of LitR, an adenosyl B_{12}-bound light-sensitive regulator of *Bacillus megaterium* QM B1151, in regulation of carotenoid production. J Bacteriol. 2015;197:2301–2315. http://dx.doi.org/10.1128/jb.02528-14

28. Lioliou EE, Pantazaki AA, Kiriakidis DA. *Thermus thermophilus* genome analysis; benefits and implications. Microbial Cell Factories. 2004;3:5–7. http://dx.doi.org/10.1186/1475-2859-3-5

29. Henne A, Brüggeman H, Raasch C, Wiezer A, Hartsch T, Liesegang H, et al. The genome sequence of the extreme thermophile *Thermus thermophilus*. Nat Biotechnol. 2004;22:547–553. http://dx.doi.org/10.1038/nbt956

30. Brey RM, Banner CDB, Wolf JB. Cloning of multiple genes involved with cobalamin (vitamin B_{12}) biosynthesis in *Bacillus megaterium*. J Bacteriol. 1986;167:623–630.

31. Ludwig ML, Drennan CL, Matthews RT. The reactivity of B_{12} cofactors: The proteins make a difference. Structure. 1996;4:505–512. http://dx.doi.org/10.1016/s0969-2126(96)00056-1

32. Ludwig ML, Matthews RT. Structure-based perspectives on B_{12}-dependent enzymes. Annu Rev Biochem. 1997;66:269–313. http://dx.doi.org/10.1146/annurev.biochem.66.1.269

33. Cheng Z, Li K, Hammad LA, Karty JA, Bauer CE. Vitamin B_{12} regulates photosystem gene expression via the CrtJ antirepressor AerR in *Rhodobacter capsulatus*. Mol Microbiol. 2014;91:649–664. http://dx.doi.org/10.1111/mmi.12491

34. Tichi MA, Tabita FR. Interactive control of *Rhodobacter capsulatus* redox-balancing systems during phototrophic metabolism. J Bacteriol. 2001;183:6344–6354. http://dx.doi.org/10.1128/jb.183.21.6344-6354.2001

35. Bauer C. Regulation of photosystem synthesis in *Rhodobacter capsulatus.* Photosynth Res. 2004;80:353–360. http://dx.doi.org/10.1023/b:pres.0000030440.99968.68

36. Randaccio L, Geremia G, Nardin G, Würges J. X-ray structural chemistry of cobalamins. Coord Chem Rev. 2006;250:1332–1350. http://dx.doi.org/10.1016/j.ccr.2005.12.005

37. Gruber TM, Gross CA. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol. 2003;57:441–446. http://dx.doi.org/10.1146/annurev.micro.57.030502.090913

38. Wikipedia contributors. Promoter (genetics) [Wikipedia page]. No date [updated 2016 Feb 07; cited 2016 Mar 02]. Available from: https://en.wikipedia.org/w/index.php?title=Promoter_(genetics)&oldid=703782251

39. Garr CD, Finke RG. Adocobalamin (AdoCbl or coenzyme B_{12}) Co-C bond homolysis radical-cage effects: Product, kinetic, mechanistic, and cage efficiency factor (F_c) studies, plus the possibility that coenzyme B_{12}-dependent enzymes function as "ultimate radical cages" and "ultimate radical traps". Inorg Chem. 1993;32:4414–4421. http://dx.doi.org/10.1021/ic00072a042

40. Garr CD, Finke RG. Radical cage effects in adocobinamide (axial-base-off coenzyme B_{12}): A simple method for trapping [Ado$^{\cdot}$ Co^{II}] radical pairs, a new β-H elimination product from the radical pair and measurement of an unprecedentedly large cage-recombination efficiency factor $F_c \geq 0.94$. J Am Chem Soc. 1992;114:10440–10445. http://dx.doi.org/10.1021/ja00052a045

41. Finke RG. Coenzyme B_{12}-based chemical precedent for Co-C bond homolysis and other key elementary steps. In: Kräutler B, Arigoni D, Golding BT, editors. Vitamin B_{12} and B_{12}-proteins. Weinheim: Wiley-VCH; 1998. p. 383–402. http://dx.doi.org/10.1002/9783527612192.ch25

42. Hay BP, Finke RG. Thermolysis of the Co-C bond of adenosylcobalamin. 2: Products, kinetics, and Co-C bond dissociation energy in aqueous solution. J Am Chem Soc. 1986;108:4820–4829. http://dx.doi.org/10.1021/ja00276a020

43. Hay BP, Finke RG. Thermolysis of the Co-C bond in adenosylcorrins. 3: Quantification of the axial base effect in adenosylcobalamin by the synthesis and thermolysis of axial base-free adenosylcobinamide. Insights into the energetics of enzyme-assisted cobalt-carbon bond homolysis. J Am Chem Soc. 1987;109:8012–8018. http://dx.doi.org/10.1021/ja00260a011

44. Schrauzer GN, Holland RJ. Hydridocobalamin and a new synthesis of organocobalt derivatives of vitamin B_{12}. J Am Chem Soc. 1971;93:4060–4062. http://dx.doi.org/10.1021/ja00745a048

45. Sension RJ, Cole AG, Harris AD, Fox CC, Woodbury NW, Lin S, et al. Photolysis and recombination of adenosylcobalamin bound to glutamate mutase. J Am Chem Soc. 2004;126:1598–1599. http://dx.doi.org/10.1021/ja0396910

46. Sension RJ, Harris AD, Stickrath A, Cole AG, Fox CC, Marsh ENG. Time-resolved measurements of the photolysis and recombination of adenosylcobalamin bound to glutamate mutase. J Phys Chem B. 2005;109:18146–18152. http://dx.doi.org/10.1021/jp052492d

47. Jarrett JT, Drennan CL, Amaratunga M, Scholten JD, Ludwig ML, Matthews RG. A protein radical cage slows photolysis of methylcobalamin in methionine synthase from *Escherichia coli*. Bioorg Med Chem. 1996;4:1237–1246. http://dx.doi.org/10.1016/0968-0896(96)00119-8

48. Baldwin DA, Betterton EA, Chemaly SM, Pratt JM. The chemistry of vitamin B_{12}. Part 25. Mechanism of the β-elimination of olefins from alkylcorrinoids; evidence for an initial homolytic fission of the Co-C bond. J Chem Soc Dalton Trans. 1985:1613–1618. http://dx.doi.org/10.1039/dt9850001613

49. Kim S-H, Chen HL, Feilchenfeld N, Halpern J. Thermal decomposition and cobalt-carbon bond dissociation energies of organocobalamins: Neopentyl-, (cyclopentylmethyl)-, (cyclohexylmethyl)-, (tetrahydrofurfuryl)- and ((tetrahydro-2*H*-pyryl)methyl)cobalamin. J Am Chem Soc. 1988;110:3120–3126. http://dx.doi.org/10.1021/ja00218a021

50. Pratt JM. Coordination chemistry of the B_{12} dependent isomerase reactions. In: Dolphin D, editor. B12. Volume 1. New York: Wiley; 1982. p. 326–392.

51. Lexa D, Savéant J-M. Brönsted basicity of vitamin B_{12}s. J Chem Soc Chem Comm. 1975; 872–4. http://dx.doi.org/10.1039/c39750000872

52. Chemaly SM, Pratt JM, The chemistry of vitamin B_{12}. Part 24. Evidence for hydride complexes of cobalt(III) corrinoids. J Chem Soc Dalton Trans. 1984:595–599. http://dx.doi.org/10.1039/dt9840000595

53. Pratt JM. The roles of Co, corrin and protein. I. Co-ligand bonding and the trans effect. In: Banerjee R, editor. Chemistry and biochemistry of B_{12}. New York: Wiley; 1999. p. 73–112.

54. Schrauzer GN, Holland RJ. Hydridocobaloximes. J Am Chem Soc. 1971;93:1505–1506. http://dx.doi.org/10.1021/ja00735a040

55. Lacy DC, Roberts GM, Peters JC. The cobalt hydride that never was: Revisiting Schrauzer's "hydridocobaloxime". J Am Chem Soc. 2015;137:4860–4864. http://dx.doi.org/10.1021/jacs.5b01838

56. Kutta RJ, Hardman SJO, Johannissen LO, Bellina B, Messiha HL, Ortiz-Guerrero, et al. The photochemical mechanism of a B_{12}-dependent photoreceptor protein. Nat Commun. 2015;6:7907–7918. http://dx.doi.org/10.1038/ncomms8907

57. SciFinder. American Chemical Society (ACS) [homepage on the Internet]. No date [cited 2015 Nov 30]. Available from: https://scifinder.cas.org/scifinder/view/scifinder/scifinderExplore.jsf

58. Kochi JK. Organometallic mechanisms and catalysis. New York: Academic Press; 1978. p. 246. http://dx.doi.org/10.1016/b978-0-12-418250-9.50017-6

59. Marques HM, Marsh JH, Mellor JR, Munro OQ. The coordination of imidazole and its derivatives by aquocobalamin. Inorg Chim Acta. 1990;170:259–269. http://dx.doi.org/10.1016/s0020-1693(00)80484-3

60. Walker LA, Jarrett JT, Anderson NA, Pullen SH, Matthews RG, Sension RJ. Time-resolved spectroscopic studies of B_{12} coenzymes: The identification of a metastable cob(III)alamin photoproduct in the photolysis of methylcobalamin. J Am Chem Soc. 1998;120:3597–3603. http://dx.doi.org/10.1021/ja974024q

61. Rury AS, Wiley TE, Sension RJ. Energy cascades, excited state dynamics, and photochemistry in cob(III)alamins and ferric porphyrins. Acc Chem Res. 2015;48:860–867. http://dx.doi.org/10.1021/ar5004016

62. Lodowski P, Jaworska M, Andruniów T, Garabato BD, Koslowski PM. Mechanism of Co-C photolysis in the base-on form of methylcobalamin. J Phys Chem A. 2014;118:11718–11734. http://dx.doi.org/10.1021/jp508513p

Reducing substance use and sexual risk behaviour among men who have sex with men in South Africa

AUTHORS:
Petal Petersen Williams[1,2]
Tara Carney[1,2]
Charles D.H. Parry[1,3]

AFFILIATIONS:
[1]Alcohol, Tobacco and Other Drug Research Unit, Medical Research Council, Cape Town, South Africa
[2]Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa
[3]Department of Psychiatry, University of Stellenbosch, Stellenbosch, South Africa

CORRESPONDENCE TO:
Petal Petersen Williams

EMAIL:
petal.petersen@mrc.ac.za

DATES:
Received: 12 Nov. 2015
Revised: 30 Mar. 2016
Accepted: 02 Apr. 2016

KEYWORDS:
HIV risk; HIV intervention; alcohol; drugs

HOW TO CITE:
Petersen Williams P, Carney T, Parry CDH. Reducing substance use and sexual risk behaviour among men who have sex with men in South Africa. S Afr J Sci. 2016;112(9/10), Art. #2015-0425, 5 pages. http://dx.doi.org/10.17159/sajs.2016/20150425

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
US Centers of Disease Control and Prevention (CDC): 5U2GPS001137-05

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

Men who have sex with men have been identified as a population at risk of acquiring and transmitting HIV. Studies in South Africa have reported a high prevalence of HIV, as well as high levels of alcohol and other drug use, among men who have sex with men, and the use of substances (alcohol and drugs) to facilitate their sexual encounters. Since 2007, interventions focused on prevention have been rolled out to vulnerable men who have sex with men and who also use alcohol or other drugs. The interventions include community-based outreach; provision of information on HIV/AIDS, substance abuse, and safer sex practices; and the development of risk-reduction plans. Among 195 men who participated in our study, there were significant reductions in the proportion who used cannabis and ecstasy, including the use of these drugs during sex. No reduction was observed in the use of any other substances. In general, after the intervention our participants reported less frequent use of alcohol and drugs and greater engagement in safer sexual practices. Despite these encouraging findings, the combination of substance use while engaging in sex had actually increased. The study findings suggest that interventions that target men who have sex with men, and who use alcohol and other drugs, could reduce risk behaviours in this population.

Significance:
- Contributes to knowledge about risk reduction strategies.
- Describes strategies for reducing drug and sexual harm among men who have sex with men.

Introduction

South Africa is a country considered to have the worst HIV epidemic worldwide. In 2012, an estimated 12.2% (6.4 million persons) of the population were HIV-positive.[1] Although researchers have not determined the national HIV prevalence among men who have sex with men (MSM) in South Africa, this group has been identified as being at particular risk of acquiring HIV. Several small-scale studies conducted among MSM have reported HIV prevalence rates of between 10.4% and 43.6%.[2-5]

Local research has also documented high levels of drug use among MSM, and the use of drugs to facilitate sexual encounters.[6] The link between drug use and HIV risk behaviour among MSM (including multiple sexual partners, unprotected anal intercourse, condom use and transactional sex) has been established in international literature.[7] The need to develop and accelerate the rollout of evidence-based interventions to address the related risks in this population has been clearly articulated.[7] South African studies have also identified the need for targeted risk reduction interventions to address the link between substance use and HIV risk behaviour among MSM.[8,9] Interventions focusing on risk reduction strategies among substance-using MSM have been shown to be successful in reducing risky sexual behavior.[10,11]

MSM in South Africa remain under-served in terms of HIV prevention and treatment services, despite studies revealing high risk of HIV infection among this population. In addition to the incidental use of drugs during sexual encounters, drugs are also used specifically to enhance sexual interactions. Aspects of such usage include enhancing the sexual experience, increasing sexual arousal, facilitating sexual encounters, increasing the capacity to engage in particular sexual activities, increasing the length of sexual interactions and facilitating sex work.[12] According to McIntyre et al.[12], despite drug use being well documented among MSM in South Africa, there are few programmes that target this aspect of HIV risk. McIntyre et al. argue that a range of possible interventions already exist and could be developed to target drug-using MSM. These include a specific focus on drug use in the context of sex parties, group sex and sex-on-site venues; non-judgemental materials focusing on risk reduction when using drugs; information about the risks of combining different substances (such as 'poppers' – amyl nitrate or butyl nitrate – and erectile dysfunction drugs); and the development of MSM-specific drug use risk reduction counselling training for health workers and people working with drug users.[12]

South Africa's second National Drug Master Plan gave prominence to the need to address drug abuse as part of broader HIV prevention efforts.[13] However, interventions targeting MSM to address drug use and sexual risk behaviour remain scarce in South Africa. Recently, the feasibility of delivering large-scale interventions to this particular group has been demonstrated locally.[14] The aim of our study was to test whether an intervention aimed at MSM who use substances (alcohol and other drugs) could affect risky substance use and sexual behaviour.

Method

In 2007, in collaboration with a local NGO in Pretoria (OUT LGBT), an initiative was begun to implement a number of harm-reduction strategies for MSM who use alcohol and other drugs. This NGO was selected because of its expertise and extensive experience working with the target population. Important components of the intervention included community outreach, distribution of condoms and lubricants, HIV risk assessment and risk reduction counselling, expanded access to HIV counselling and testing, care and treatment of HIV and sexually transmitted

diseases, and referrals to substance abuse treatment and other social services. The intervention also aimed to build referral networks, decrease stigma among service providers, and influence NGO practices and government policies on drugs and HIV. These activities were in line with the six core elements defined in the *Technical Guidance on Combination HIV Prevention* for effective MSM programmes as released by the Office of the U.S. Global AIDS Coordinator in May 2011, as part of the U.S. President's Emergency Plan for AIDS Relief's overall prevention strategy.[15]

Participants and data collection

The intervention was conducted in various areas in and around Johannesburg and Pretoria in Gauteng, and around Nelspruit in Mpumalanga. A number of locations were targeted, including gay and lesbian pride events, gay clubs and areas known to be frequented by MSM, and through site work at the NGO's premises or clinics and wellness centres, where the NGO made use of negotiated spaces within such clinics or centres.

To be included in the study, participants had to be 16 years or older. They also had to be self-reported MSM who were using alcohol and/or drugs, regardless of whether they self-identified as gay, bisexual or heterosexual, and regardless of HIV status. The quantity of alcohol or drugs used by participants was not set as an exclusion criterion. Participants who were eligible for inclusion and willing to participate signed an informed consent form, and the principles of the Declaration of Helsinki were adhered to. Peer outreach workers completed a face-to-face baseline questionnaire with participants, to record risk behaviours. A risk-reduction plan was developed with each participant, which took into account risks related to injection and non-injection drug use, sex-related risks and HIV testing.

Intervals between follow-up appointments were not specified. In some instances outreach workers made appointments for follow-up at a time that suited the participants, and in other instances follow-up appointments occurred spontaneously when outreach workers met up with participants at various events or social gatherings. At follow-up, the same questionnaire that had been administered at baseline was completed again. Also at follow-up, clients were asked the frequency of risk behaviour or its reduction, and whether they had been tested for HIV or asked their partner to test for HIV or encouraged a friend to be tested. Behaviour change was thus self-reported rather than observed.

In the areas where the intervention was delivered, MSM participants spoke Sepedi, Setswana, Afrikaans or English. Questionnaires were administered in English, with outreach workers translating difficult concepts where necessary. Our study is part of a bigger study that includes four provinces of South Africa. Ethical approval for conducting the study was granted by the Health Research Committee of the University of Stellenbosch.

Intervention

The NGO received training on an intervention that was based on a local adaptation of the World Health Organization's *Training Guide for HIV Prevention Outreach to Injecting Drug Users*.[16] The adaptation lessened the focus on injection drug use related behaviours, and placed greater focus on substance-related sexual HIV risk behaviour. The adapted manual also emphasised drugs commonly used in South Africa. The NGO recruited peer outreach workers on a volunteer basis and they were paid a modest stipend. Project coordinators were appointed in the organisation to ensure the smooth running of the project, and to liaise between the NGO and the Medical Research Council project manager.

The intervention included community-based outreach and provision of information on HIV/AIDS, substance use, and safer sex practices. Each MSM who participated developed a personalised HIV risk reduction plan with the help of the outreach workers. The plans were followed up and reassessed with the participants. In addition to the provision of risk reduction counselling, the intervention included the provision of HIV counselling and testing services and referrals to treatment and care for substance abuse, HIV or sexually transmitted infections, and other social services. Intervention services were monitored on a monthly basis and evaluated biannually.

Data analysis

Demographic information was collected at the first contact (baseline), and descriptive statistics were calculated for these variables. Data on substance use and sexual risk behaviour were also collected at baseline and at each follow-up appointment. We assessed the distribution of behavioural data using the Shapiro-Wilks test of normality, and the results indicated that the data were non-normally distributed ($p<0.001$). Therefore, we conducted bivariate analysis for continuous variables (number of sex partners, number of times engaged in sex, number of times had unprotected sex, number of times traded sex, and number of times had sex under the influence of substances). The Wilcoxon signed rank test was used to determine whether self-reported differences between baseline and follow-up were statistically significant. For categorical variables (types of substances used, either in general or during sex), we assessed the differences in proportions with the chi-square test of association. All statistics were analysed at 95% confidence intervals, and data analysis was performed with SPSS version 21.

Results

Sample characteristics

The early years of implementation of the intervention were 2007 to 2009. This paper reports on implementation activities between 2010 and 2012. Baseline information was collected from 195 MSM who were recruited into the study, whose substance abuse and HIV risk profile was assessed. The participants then received an intervention and were subsequently followed up. Of these, 27.7% had one repeat contact with outreach workers, 30.3% had two repeat contacts, 29.7% had three repeat contacts and 10.8% had four repeat contacts. The data presented are for first contact (baseline) and the final follow-up contact for a particular participant. The median age of the MSM participants was 27 years, and participants had a median of 12 years of formal education. The majority of participants in our study were employed (56.4%) and single (66.7%). Table 1 shows the demographic data.

Table 1: Summary statistics of participants' demographic data ($N=195$)

	N	**%**
Occupation		
Unemployed	41	21.0
Employed	110	56.4
Student or pupil	39	20.0
Other	4	2.1
Marital Status		
Single	130	66.7
Partnered	53	27.2
Married, opposite sex	7	3.6
Married, same sex	1	0.5
Divorced or separated	3	1.5
Age (years)	27 (median)	31 (18–49)[a]
Education (years)	12 (median)	6 (7–13)[a]

[a] *Interquartile range*

Substance use

We calculated the proportion of participants who used each substance at baseline (time$_1$) and at last follow-up (time$_2$), and the difference between these two time points (Table 2). Substances used by less than 5% of the sample were excluded. Only four of the 195 participants reported injection drug use (at baseline). There were significant reductions in the proportion of MSM who used cannabis ($p=0.009$) and ecstasy ($p=0.043$) from time$_1$ to time$_2$. No significant differences were observed for the use of alcohol, cocaine, inhalant, amyl or butyl nitrate, and over-the-counter or prescription drugs from time$_1$ to time$_2$.

Table 2: Prevalence of substance use at time$_1$ and time$_2$

Substance	*N* (time$_1$)	*N* (time$_2$)	*z*	*p*
Alcohol	193	191	0.81 (-0.014;0.034)	0.417
Cannabis	126	101	2.61 (0.03;0.227)	0.009*
Cocaine	14	12	0.40 (-0.039;0.059)	0.689
Ecstasy	33	19	2.02 (0.003;0.137)	0.043*
Inhalants	29	25	0.57 (-0.049;0.089)	0.569
Amyl or butyl nitrate	30	32	0.27 (-0.08;0.06)	0.785
Over-the-counter or prescription drugs	10	3	1.61 (-0.006;0.066)	0.107

There was also a highly significant change in the frequency of substance use ($p\leq0.001$), with participants using substances less frequently after the intervention. At time$_1$, 30.3% reported daily use of substances, whereas only 13.8% reported daily use at time$_2$, with a greater number of participants using substances only once a week or less (Table 3).

Table 3: Frequency of substance use at time$_1$ and time$_2$

Frequency of use	time$_1$	time$_2$
Once a week or less	16.4%	31.8%
2-6 days a week	50.8%	54.4%
Daily	30.3%	13.8%

The use of more than one substance (polydrug use) was recorded at time$_1$ and time$_2$. After the intervention, 35.9% of participants reported no change in the number of different substances used, and 28.2% had actually increased the number of different substances they used. However, 35.9% had decreased the number of different substances used between time$_1$ and time$_2$ (Table 4).

Table 4: Change in drug and alcohol use

Type of change	*N*	%
No change	70	35.9
Increase in AODs by number of substances:		
1	44	22.6
2	8	4.1
3	2	1.0
4	1	0.5
Decrease in AODs by number of substances:		
1	35	17.9
2	30	15.4
3	5	2.6
4	0	0.0

Key: AOD, alcohol and other drugs

Sexual risk behaviour

Following the intervention, participants had significantly fewer partners ($z=-6.663$; $p\leq0.001$) but had sex with these partners more frequently (Table 5). Frequency of receptive anal sex increased significantly ($z=-5.551$; $p\leq0.001$); frequency of insertive anal sex increased significantly ($z=-2.707$; $p=0.007$) and frequency of oral sex also increased significantly ($z=-5.409$; $p\leq0.001$). However, despite having sex more frequently, the participants reported having safer sex following the intervention. The frequency of condom usage increased significantly for receptive anal sex ($z=-7.960$; $p\leq0.001$), insertive anal sex ($z=-6.094$; $p\leq0.001$) and oral sex ($z=-4.411$; $p\leq0.001$).

Table 5: Sexual risk behaviour reported at time$_1$ and time$_2$

Sexual risk	Time$_1$ Median (IQR)%	Time$_2$ Median (IQR)%	*z*	*p*
Number of sex partners	3 (1-15)	2 (1-17)	-6.663	0.000*
Times had receptive anal sex	6 (0-31)	10 (0-45)	-5.551	0.000*
Times had insertive anal sex	8 (0-40)	10 (0-40)	-2.707	0.010*
Times had oral sex	10 (0-60)	19 (0-60)	-5.409	0.000*
Times used condoms for receptive anal sex	1 (0-22)	9 (0-45)	-7.960	0.000*
Times used condoms for insertive anal sex	3 (0-26)	7 (0-38)	-6.094	0.000*
Times used condoms for oral sex	0 (0-7)	0 (0-26)	-4.411	0.000*
Times traded sex for money	0 (0-25)	0 (0-39)	-2.005	0.050

**, $p<0.05$*

Key: IQR, interquartile range

Sexual risk and substance use

As shown in Table 6, there were significant reductions in the proportion of participants who used cannabis ($p=0.005$) and ecstasy ($p=0.04$) during sex, from time$_1$ to time$_2$. No significant differences were observed between time$_1$ and time$_2$ for the use of alcohol, cocaine, inhalants, amyl or butyl nitrate, and over-the-counter or prescription drugs during sex. However, participants engaged in sex while using drugs and alcohol more frequently after the intervention (median sexual events: time$_1=10$; time$_2=18$; $z=-3.465$; $p=0.001$). In total, 34.4% of participants reported no change in the number of different substances used during sex, whereas 28.7% had increased the number of different substances used during sex, and 36.9% had decreased the number of different substances used during sex (Table 7).

Table 6: Sexual risk and substance use reported at time$_1$ and time$_2$

Substance	*N* (Time$_1$)	*N* (Time$_2$)
Alcohol during sex	189	187
Cannabis during sex	129	101
Cocaine during sex	15	10
Ecstasy during sex	32	18
Inhalants during sex	28	25
Amyl/butyl nitrate during sex	27	31
OTC/Pre during sex	9	3

Table 7: Change in alcohol and drug use during sex

Type of change	*N*	%
No change	67	34.4
Increase in AODs by number of substances:		
1	45	23.1
2	9	4.6
3	1	0.5
4	1	0.5
Decrease in AODs by number of substances:		
1	34	17.4
2	32	16.4
3	6	3.1
4	0	0.0

Key: AODs, alcohol and other drugs

Discussion

MSM participants in our study engaged in multiple alcohol or other drug use and sexual risk behaviours. However, as reported in an earlier phase of this project[14], participants were receptive to NGOs delivering HIV prevention and substance risk reduction services, and were willing to develop individual risk reduction strategies with outreach workers. Other intervention studies with substance-using MSM have also demonstrated that it is possible to enrol and retain MSM in intervention studies.[7]

Our study showed that the only significant reduction after the intervention was the proportion of MSM who used cannabis and ecstasy, including the use of these drugs during sex. However, participants did report *less frequent* alcohol or other drug use at $Time_2$. Additionally, more participants had reduced the diversity of substances used relative to the number of participants who increased their polysubstance use (including the use of substances during sex). Most importantly, participants reported safer sex practices following the intervention, including having fewer sexual partners and more frequent condom use during sex. However, after the intervention, participants engaged in sex while using drugs and alcohol more frequently than they had before the intervention. Other studies have similarly found that risk reduction, rather than 'abstinence only' strategies, are an effective approach for this population.[17]

Evidence suggests that behavioural interventions to reduce risk for sexual transmission of HIV among MSM can be effective. A meta-analysis that reviewed HIV behavioural interventions for reducing sexual risk behaviour of MSM found, as did our study, that interventions were associated with a significant decrease in unprotected anal intercourse and number of sexual partners, and with a significant increase in condom use during anal intercourse.[11] A Cochrane review found that these interventions can lead to significant risk reduction among MSM, specifically a reduction in unprotected anal intercourse.[18] However, the authors distinguish between individual, group and community-level interventions; not all interventions are equally effective.[18]

A randomised controlled trial was conducted in the U.S. to test a cognitive-behavioural group intervention outside treatment centres, focusing on reducing sexual risk behaviour of substance-using MSM.[19] The study found the intervention was no more efficacious than the control condition in reducing sexual risk behaviours or substance use. Instead, substantial reductions occurred in both the intervention and control groups.[19] In another trial examining the effects of motivational interviews on club drug use and risky sex, with MSM participants who did not seek treatment, motivational interviews were not found to be effective in reducing risky sexual behaviour.[20] However, the motivational interviews did result in significant reductions in club drug use in the intervention group compared with the control group – but only among participants with lower severity of drug dependence.[20] Based on these findings, it appears that although risk-reduction interventions have demonstrated their ability to reduce risky behaviour, further research is needed to determine which interventions are most efficacious.[7]

The findings of our study are subject to the following limitations. Firstly, risk behaviours were self-reported by the MSM participants and might be subject to social desirability or other biases. Although peer outreach workers established relationships with the participants and this may have facilitated honest self-reporting, we did not confirm the self-reports by biological testing for substance use. Secondly, questionnaires were administered in English, with outreach workers translating difficult concepts where necessary. This approach is not as robust as written and tested translation, and interpretation errors can occur when English-medium tools are used with non-native English speakers. Thirdly, there was no comparison group and all participants received the same intervention. Thus, it was not possible to conclude that changes in risky behaviour were the result of the intervention. Finally, we recruited participants only from certain areas in and around Gauteng and Mpumalanga, and the results should not be generalized to other provinces and areas. As reported in other studies, our participants were recruited by outreach workers from local MSM NGOs, which further limits the generalizability of the findings.[21]

Despite these limitations, the findings of our study (and others) demonstrate that interventions targeting substance-using MSM can reduce risk behaviour among this population. In South Africa, programmes that target MSM for HIV prevention, treatment and care are not well developed. Where they do occur, they are provided mostly by NGOs or community-based organisations in specific geographical areas and have not been up-scaled across the country.[12] As a result there remains a need to offer widespread integrated services related to alcohol or drug use and HIV risk, especially where community-based outreach is available and is tailored for MSM who use alcohol or other drugs. What the current study has shown is that while interventions with MSM can have a positive effect in reducing HIV risk behaviours (such as having sex without a condom), the effects on substance-use behaviour are modest. More work is needed to determine how substance abuse behaviour among MSM can be reduced effectively.

Authors' contributions

C.P. designed the study and was the project leader. P.P.W. and T.C. performed the analysis, and P.P.W. wrote the first draft of the manuscript. All authors contributed to and approved the final manuscript.

References

1. Shisana O, Rehle T, Simbayi LC, Zuma K, Jooste S, Zungu N, et al. South African national HIV prevalence, incidence and behaviour survey, 2012. Cape Town: HSRC Press; 2014.
2. Burrell E, Baral S, Beyrer C, Wood R, Bekker LG. Abstract LBPE1195. IAS Conference 2008; Mexico City, Mexico, 3–8 August 2008.
3. Burrell E, Mark D, Grant R, Wood R, Bekker LG. Sexual risk behaviours and HIV-1 prevalence among urban men who have sex with men in Cape Town, South Africa. Sex Health. 2010;7:149–153. http://dx.doi.org/10.1071/SH09090
4. Lane T, Fisher Raymond H, Dladla S, Rasethe J, Struthers H, McFarland W, et al. HIV prevalence among men who have sex with men in Soweto, South Africa: Results from the Soweto men's study. AIDS Behav. 2011;15:626–634. http://dx.doi.org/10.1007/s10461-009-9598-y
5. Metcalf C, Rispel L. Exposing a hidden HIV epidemic among men who have sex with men. HSRC Review. 2009;7(2):4–5. http://dx.doi.org/10.1016/S0968-8080(09)33442-4
6. Parry CDH, Petersen P, Dewing S, Carney T, Needle R, Kroeger K, et al. Rapid assessment of drug related HIV risk among men who have sex with men in three South African cities. Drug Alcohol Depen. 2008;95:45–53. http://dx.doi.org/10.1016/j.drugalcdep.2007.12.005
7. Santos GM, Das M, Colfax GN. Interventions for non-injection substance use among men who have sex with men: What is needed. AIDS Behavior. 2011;15:S51–S56. http://dx.doi.org/10.1007/s10461-011-9923-0

8. Lane T, Shade SB, McIntyre J, Morin SF. Alcohol and sexual risk behavior among men who have sex with men in South African township communities. AIDS Behavior, 2008;12(1):78–85 http://dx.doi.org/10.1007/s10461-008-9389-x

9. Parry CDH, Petersen P, Carney T, Needle R. Opportunities for enhancing and integrating HIV and drug services for vulnerable drug using populations in South Africa. Int J Drug Policy. 2010;21(4):289–295. http://dx.doi.org/10.1016/j.drugpo.2009.11.008

10. Mausbach BT, Semple SJ, Strathdee SA, Zians J, Patterson TL. Efficacy of a behavioral intervention for increasing safer sex behaviors in HIV-positive MSM methamphetamine users: Results from the EDGE study. Drug Alcohol Depen. 2007;87(2–3):249–257. http://dx.doi.org/10.1016/j.drugalcdep.2006.08.026

11. Herbst JH, Sherba RT, Crepaz N, DeLuca JB, Zohrabyan L, Stall RD, et al. A meta-analytic review of HIV behavioral interventions for reducing sexual risk behavior of men who have sex with men. J Acq Immun Def Synd. 2005;39(2):228–241.

12. McIntyre J, Jobson G, Struthers H, De Swardt G, Rebe K. Rapid assessment of HIV prevention, care and treatment programming for MSM in South Africa. Assessment Report 2013. Johannesburg: Anova Health Institute. http://dx.doi.org/10.1007/s10461-012-0328-5

13. Department of Social Development. National drug master plan 2006–2011. Republic of South Africa, Pretoria: Department of Social Development; 2007.

14. Petersen Williams P, Carney T, Parry CDH, Plüddemann A. Intervening to identify and reduce drug use and sexual HIV risk patterns among men who have sex with men in three provinces in South Africa. J Subst Use. 2014;19(1–2):141–146. http://dx.doi.org/10.3109/14659891.2012.760009

15. The U.S. President's Emergency Plan for AIDS Relief. Technical guidance on combination HIV prevention. [Document on the Internet.] c 2011 [Cited 2011 September 14]. Available from http://www.pepfar.gov/documents/organization/164010.pdf.

16. Burrows D. Training guide for HIV prevention outreach to injecting drug users: Workshop manual. Geneva: World Health Organization, Department of HIV/AIDS; 2004.

17. Seal DW, Kelly JA, Bloom FR, Stevenson LY, Coley BI, Broyles LA, et al. HIV prevention with young men who have sex with men: What young men themselves say is needed. AIDS Care. 2000;12(1):5–26. http://dx.doi.org/10.1080/09540120047431

18. Johnson WD, Diaz RM, Flanders WD, Goodman M, Hill AN, Holtgrave D, et al. Behavioral interventions to reduce risk for sexual transmission of HIV among men who have sex with men. Cochrane DB Syst Rev. 2008;3, Art. #CD001230. http://dx.doi.org/10.1002/14651858.CD001230.pub2

19. Mansergh G, Koblin BA, McKirnan DJ, Hudson SM, Flores SA, Wiegand RE, et al. An intervention to reduce HIV risk behavior of substance-using men who have sex with men: A two-group randomized trial with a nonrandomized third group. PLoS Med. 2010;7(8), Art. #e1000329. http://dx.doi.org/10.1371/journal.pmed.1000329

20. Morgenstern J, Bux DA Jr, Parsons J, Hagman BT, Wainberg M, Irwin T. Randomized trial to reduce club drug use and HIV risk behaviors among men who have sex with men. J Consult Clin Psychol. 2009;77(4):645–656. http://dx.doi.org/10.1037/a0015588

21. Thomas B, Mimiaga MJ, Mayer KH, Johnson CV, Menon S, Chandrasekaran V, et al. HIV prevention interventions in Chennai, India: Are men who have sex with men being reached? AIDS Patient Care ST. 2009;23(11):981–986. http://dx.doi.org/10.1089/apc.2009.0092

Do arthropod assemblages fit the grassland and savanna biomes of South Africa?

AUTHORS:
Monique Botha[1]
Stefan J. Siebert[1]
Johnnie van den Berg[1]

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

CORRESPONDENCE TO:
Monique Botha

EMAIL:
moniquebotha210@gmail.com

DATES:
Received: 12 Nov. 2015
Revised: 26 Mar. 2016
Accepted: 06 Apr. 2016

KEYWORDS:
species distribution; arthropod trophic groups; cluster analysis; biome-specificity; arthropod species composition

HOW TO CITE:
Botha M, Siebert SJ, van den Berg J. Do arthropod assemblages fit the grassland and savanna biomes of South Africa? S Afr J Sci. 2016;112(9/10), Art. #2015-0424, 10 pages. http://dx.doi.org/10.17159/sajs.2016/20150424

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
GenØk-Centre of Biosafety

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

The long-standing tradition of classifying South Africa's biogeographical area into biomes is commonly linked to vegetation structure and climate. Because arthropod communities are often governed by both these factors, it can be expected that arthropod communities would fit the biomes. To test this hypothesis, we considered how well arthropod species assemblages fit South Africa's grassy biomes. Arthropod assemblages were sampled from six localities across the grassland and savanna biomes by means of suction sampling, to determine whether the two biomes have distinctive arthropod assemblages. Arthropod samples of these biomes clustered separately in multidimensional scaling analyses. Within biomes, arthropod assemblages were more distinctive for savanna localities than grassland. Arthropod samples of the two biomes clustered together when trophic groups were considered separately, suggesting some similarity in functional assemblages. Dissimilarity was greatest between biomes for phytophagous and predacious trophic groups, with most pronounced differentiation between biomes at sub-escarpment localities. Our results indicate that different arthropod assemblages do fit the grassy biomes to some extent, but the pattern is not as clear as it is for plant species.

Significance:

- Provides the first comparison of arthropod composition between grassland and savanna biomes of South Africa.
- Explores whether these two biomes show distinct arthropod assemblages.
- Documents the characteristics of arthropod assemblages.
- Confirms that plant assemblages of biomes are more distinguishable than arthropod assemblages.

Introduction

South Africa's rich biodiversity is largely the result of a wide range of climatic conditions and topographic variation, which give rise to relatively distinctive biomes, each with characteristic plant and animal species.[1,2] Vegetation categorisation in South Africa is nested within these biome concepts.[2] Insects are particularly relevant in biome comparisons as a large proportion of insects may be host-specific phytophagous species,[3] which is likely to make them vegetation-specific because of the intricate relationships. European studies have shown that local plant species composition is the most effective predictor of arthropod assemblage composition, even more so than vegetation structure and environmental conditions.[4] Furthermore, arthropod groups have been shown to be associated with particular plant assemblages in grassland, with certain insect orders responding positively to the increase in specific plant functional groups.[5]

Being ectotherms, arthropods are sensitive to their abiotic environments. The vegetation layer provides a biotic environment that buffers arthropods against changes in the abiotic environment. Several studies have shown that factors such as vegetation height, density and percentage cover, as well as the associated microclimate, have significant effects on species composition of grasshoppers[6-8] and dung beetles[9].

Despite the proven direct and indirect relationships between plant and insect composition, little research has been conducted on specific structures of insect communities in southern African biomes.[10-12] A study of four biomes in South Africa revealed that overall, differences between insect assemblages of different biomes are not as convincing as those between plant assemblages.[11] This is to be expected, considering the better dispersal ability of most insects because of their mobility and the frequent dispersal events characteristic of winged species. The transition from one biome to another therefore appears smoother for insect assemblages than it is for plant assemblages.[11]

In a previous study of maize-producing regions in South Africa, we collected data to compare insect and plant diversity of field margin habitats in two grassy biomes.[13] In the current study, we used the earlier non-crop dataset of untransformed areas to assess the compositional similarities and differences between the assemblages of arthropods and plants from grassland and savanna biomes. This contribution provides a first comparison of the arthropod composition of localities in these biomes, to establish whether they have distinctive arthropod assemblages.

Savannas are multi-structured and therefore have great structural complexity and niche diversification to house a wide variety of arthropod species.[14] As a result, the presence of a tree and shrub layer allows for more arthropod species and higher abundances per unit area compared with habitats that have simpler structures, such as grassland.[15,16] Savanna provides a wider variety of conditions and resources to be exploited, allowing a greater degree of species coexistence.[17,18] Considering the plant species, life forms and structure of savanna, we wanted to test whether the composition of arthropod assemblages varied between the grassland and savanna biomes – especially considering that biomes are defined by dominant vegetation and climatic variables. We tested whether arthropod assemblages in the grassland and savanna biomes follow the same biogeographical patterns as plant assemblages, and whether these biomes can be differentiated by arthropod species composition.

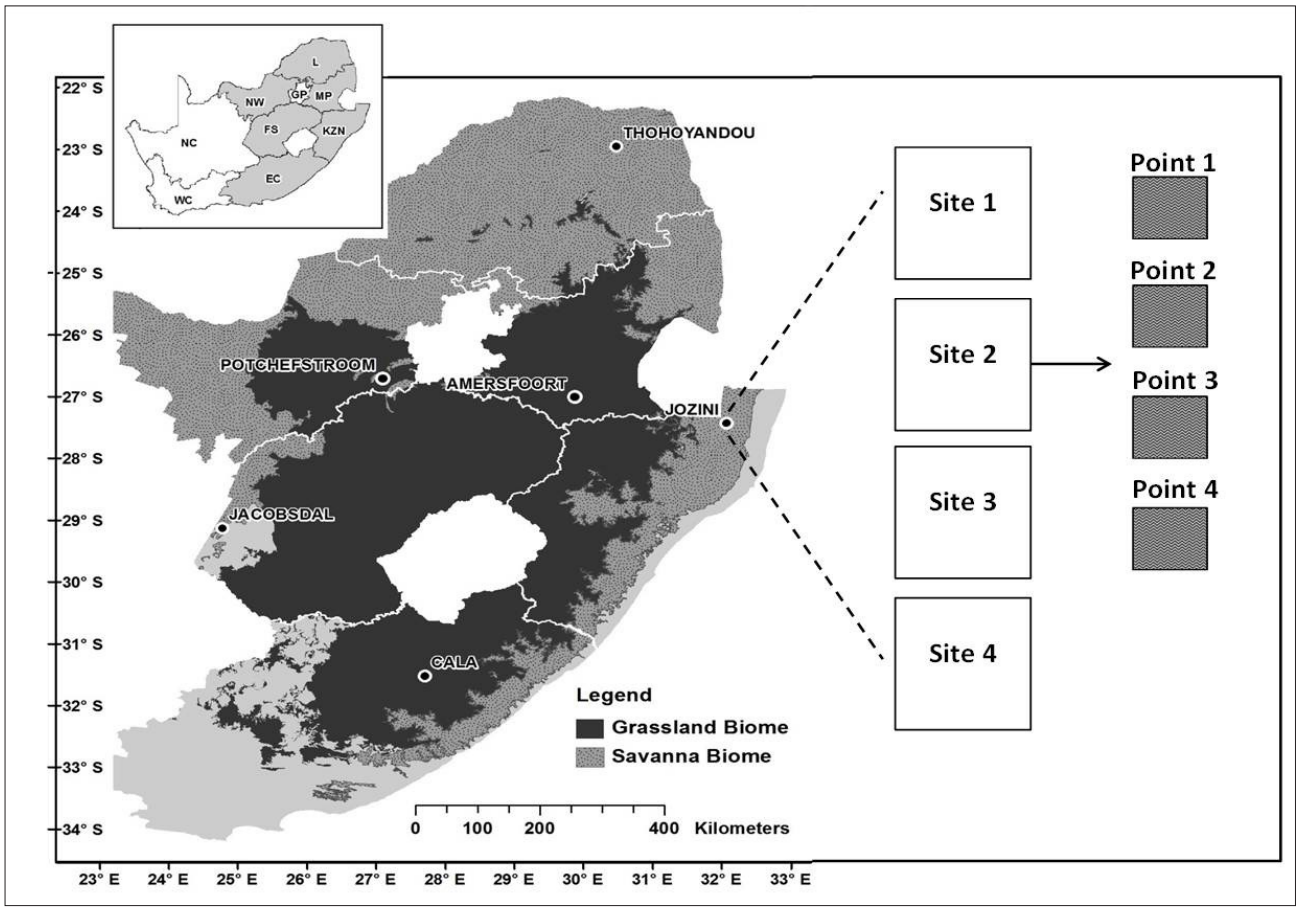


Figure 1: Localities across the six major maize-producing provinces of South Africa within grassland and savanna biomes. EC, Eastern Cape; FS, Free State; KZN, KwaZulu-Natal; L, Limpopo; MP, Mpumalanga; NW, North-West. Three localities were situated in the grassland biome (Amersfoort, Cala and Potchefstroom) and three in the savanna biome (Jacobsdal, Jozini and Thohoyandou). Within each locality, four sites were chosen as indicated. Each site contained four sampling points situated in a crop field margin habitat.

Methods

Experimental design

Our study investigated plants and plant-dwelling arthropods, sampled at six localities spread throughout South Africa in the core regions of the grassland and savanna biomes. The study covered the sub-escarpment (Cala and Jozini), escarpment (Amersfoort and Thohoyandou) and interior plateau (Potchefstroom and Jacobsdal) areas of each biome (Figure 1). To minimise possible spatial autocorrelation of data, the two localities representing similar elevations were chosen to represent different biomes (for instance, the sub-escarpment locality Cala represented grassland and the sub-escarpment locality Jozini represented savanna).

Arthropod data were generated within a total sampled area of 2400 m^2 (96 plots of 5 m x 5 m), and plant data were generated from a total sampled area of 38 400 m^2 (96 plots of 20 m x 20 m). At each locality, four sites were sampled 5 km apart, in a spatial layout designed to cover assemblage variation across several spatial scales.[13] To avoid pseudo-replication, two sites were sampled on hillslopes and two in the valleys of each locality. At each site we sampled two pairs of points 100 m apart at both ends of a 1-km transect, with one site upslope and the other downslope. This translated to 4 points per site, 16 per locality and 48 points for each of the biomes (Table 1). The data were collected during late morning and were scheduled to coincide with the season of maximum biological activity (January and February). Plants and insects that occur seasonally were therefore under-represented at some sites.

Table 1: Layout of sampling design indicating the number of sample repeats (*n*) in brackets for the respective areas

	Biome	Topographic region	Locality
Levels	Grassland (48)	Sub-escarpment (16)	Cala (16)
		Escarpment (16)	Amersfoort (16)
		Plateau (16)	Potchefstroom (16)
	Savanna (48)	Sub-escarpment (16)	Jozini (16)
		Escarpment (16)	Thohoyandou (16)
		Plateau (16)	Jacobsdal (16)

Arthropod sampling

Suction sampling of arthropods using an adapted D-Vac method[13] was conducted in 5 m x 5 m sampling points. The D-Vac provides a relatively fast method for sampling large areas of vegetation, although its effectiveness may be altered by weather conditions and vegetation characteristics. It was most effective in dry, upright grassy habitats. A further limitation was that certain species may selectively be extracted and others under-sampled during suction-sampling in dense vegetation.[19] However, compared with passive sampling methods, the D-Vac method is not as dependent on insect activity, is less prone to sampling error and

may represent one of the best techniques for sampling a wide range of arthropod taxa on vegetation.[20]

Seven swaths per plot were made, following a zigzag pattern with the D-Vac nozzle for each swath. Each swath was made through the vegetation from one side of the plot to the other. Where tall grass (>1 m), shrubs and trees were present within sampling points, arthropod individuals were sampled by moving the D-Vac over the branches and large leaves as well as the trunk or stem, up to a height of 2 m. Vegetation beneath dense shrubs and trees was also sampled where accessible. Soil-dwelling arthropods that were present on the lower parts of plants during the survey were also collected. We did not attempt to use the D-Vac to collect from the soil surface.

Vegetation sampling

After the arthropod sampling was completed, a fixed-width (2-metre) line transect approach was followed[21] to record plant species. The plot was adapted to include ten parallel transects (2 m wide and 20 m long). This ensured that the vegetation sampling would overlap and extend beyond the arthropod sampling area. At 1-metre intervals, one plant species was recorded for every major growth form – that is grass, forb, shrub and tree. Four species were therefore recorded at each interval, if all growth forms were represented. The number of individuals per species across the ten transects (100 points) was summed to determine the species abundances for each sample point.

Table 2: Number (*n*) of plant and arthropod species, and percentage of individuals per family, order or trophic group, and for each biome

		Grassland Biome		Savanna Biome		Total	
		Species (*n*)	**Individuals %**	**Species (*n*)**	**Individuals %**	**Species (*n*)**	**Individuals %**
Plant families	Poaceae	62	52.1	73	35.5	109	42.6
	Fabaceae	28	4.7	73	13.2	94	9.6
	Asteraceae	46	17.3	32	6.9	73	11.3
	Acanthaceae	7	1.4	25	7.3	30	4.8
	Rubiaceae	4	3.9	22	1.8	26	2.7
	Apocynaceae	9	0.3	16	1.5	25	1.1
	Euphorbiaceae	4	0.2	23	2.8	25	1.7
	Cyperaceae	17	3.4	7	0.9	24	1.9
	Malvaceae	6	0.6	20	1.9	22	1.4
	Lamiaceae	4	0.2	16	0.9	20	0.6
	Other families (83)	84	15.9	227	27.1	302	22.3
	All plants (93)	272	–	534	–	751	–
Arthropod orders	Hemiptera	198	25.9	194	32.3	340	26.7
	Diptera	126	8.9	128	18.7	233	10.1
	Hymenoptera	128	14.9	120	16.1	225	15.1
	Araneae	97	9.5	87	8.5	165	9.3
	Coleoptera	80	2.7	86	5.5	147	3.1
	Orthoptera	52	1.4	94	7.5	127	2.1
	Lepidoptera	37	0.6	62	3.5	89	0.9
	Acari	25	23.4	15	3.8	32	21.3
	Thysanoptera	15	4.2	15	1.3	22	3.8
	Mantodea	5	0.06	13	0.7	18	0.1
	Other orders (13)	23	8.4	23	2.2	38	7.4
	All arthropods (23)	786	–	837	–	1436	–
Arthropod trophic groups	Herbivores	339	43.1	430	51.4	667	46.4
	Predators	179	22.8	158	18.9	297	20.7
	Parasitoids	115	14.6	104	12.4	204	14.2
	Pollinators	49	6.2	75	9.0	112	7.8
	Other groups (5)	148	18.8	131	15.7	252	17.5

Statistical analysis

The non-parametric species estimators of observed species counts (Sobs), Chao2 and Jacknife1[22,23] were calculated using PRIMER 7[24] to determine how closely the sample resembled the extrapolated species richness. Non-metric multidimensional scaling (NMDS) analyses (samples clustered based on Bray–Curtis dissimilarity) in PRIMER 7 were used to visualise differences between sampling points in ordination space, in terms of plant and arthropod assemblages. For 2-dimensional ordinations, the stress value increases with decreasing dimensionality and increasing quantity of data. The general rule is as follows: stress ≤ 0.05 gives an excellent representation, with no prospect of misinterpretation of the data, and stress ≤ 0.1 represents a good ordination with no real risk of misinterpretation. Stress ≤ 0.2 may still give a potentially useful ordination, but cross-checks with other techniques are recommended.[25]

Significance of NMDS clusters were tested by permutational MANOVA (PERMANOVA), analysis of similarities (ANOSIM) and similarity percentage (SIMPER) analyses, using PRIMER 7. PERMANOVA is a multivariate analysis of variance technique suitable for abundance data and where significance is based on permutation of the dissimilarity matrix.[26] First, PERMANOVA was conducted using a Bray–Curtis dissimilarity matrix to determine the main and interactive effects of biome and topographic region on species composition (permutations = 999; type III sums of squares). Then ANOSIM was used as a post hoc test for pairwise comparisons between localities within the grassland and savanna biomes, to assess compositional dissimilarity. ANOSIM is a non-parametric test that uses rank dissimilarities based on the Bray–Curtis coefficient of similarity. Significant separation of two distinct clusters in ordinal space is calculated using an *R*-statistic, which ranges from 1 (meaning clusters are totally different) to 0 (meaning clusters are indistinguishable).[27] Next, SIMPER was applied to the data set to assess which taxa were primarily responsible for observed differences in species composition between the biomes. A square root transformation of species data was performed for NMDS, PERMANOVA, ANOSIM and SIMPER analyses to reduce the influence of common species.[26]

As a final measure, canonical correspondence analysis (CCA) with forward selection was applied to the data, using CANOCO 4.5[28], as a cross-check to depict how different localities compared in terms of arthropod and plant species composition. The same analysis enabled us to assess the relative importance of selected environmental variables in determining plant and arthropod assemblages of grouped sampling points. Five biotic and abiotic environmental factors were considered for a biplot with species data. These included latitude and longitude (decimal degrees), altitude (m.a.s.l.), tree cover (%) and grass cover (%). Species data for CCA analyses were square-root transformed, and environmental data were normalised.

Results

The survey recorded 1436 arthropod morpho-species (35 193 individuals) from 23 orders (Table 2). The four largest trophic groups of arthropods were distinguished further for comparative analyses, namely herbivores (667 morpho-species), predators (297), parasitoids (204) and pollinators (112) (Table 2). Other groups not included in the analyses were decomposers, parasites, visitors, frugivores and omnivores. For plants, 740 species (10 856 individuals) from 93 families were recorded in the field margin habitats (Table 2).

Some groups of plants and arthropods (mobile, sensitive insects such as butterflies or grasshoppers and cryptic plants such as geophytes) might be under-represented in the samples because of the collection methods and specificity of season. It must be mentioned that a study targeting soil and flying arthropods could yield very different results. A selection of species accumulation curves (Figure 2) suggested that the saturation levels were not satisfactory for arthropods, and this should be kept in mind when interpreting the findings[22].

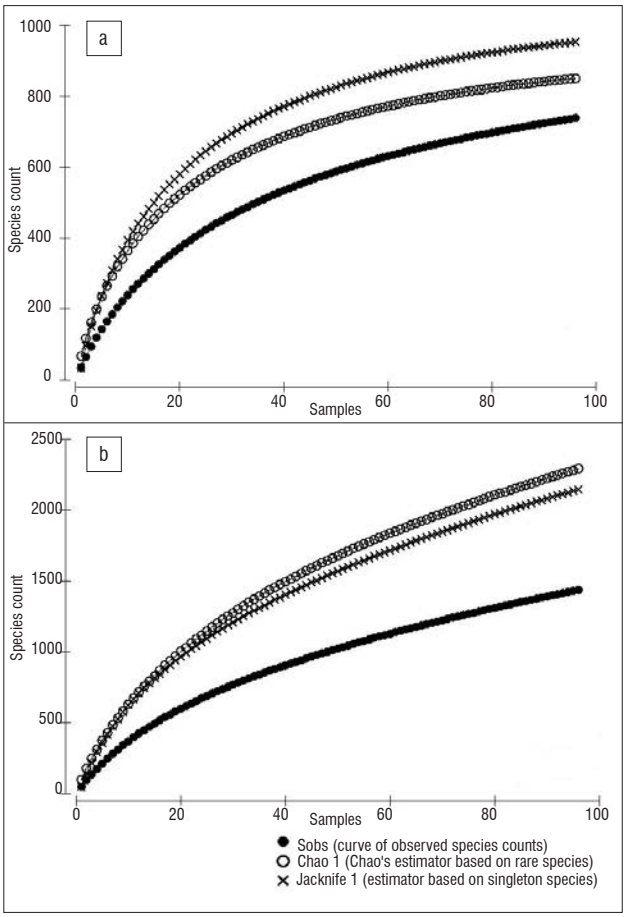


Figure 2: Species accumulation curves for (a) plants and (b) arthropods sampled in all grassland and savanna plots.

Plant assemblages across biomes

The NMDS analysis for plants revealed the tightest clustering and the lowest stress factor (0.14) compared with the results for arthropods (Figure 3). Clear distinctions in plant species composition were found between grassland and savanna sampling points, with a much tighter clustering for grassland samples (Figure 3a). Differences in plant species composition between biomes were confirmed by PERMANOVA results (pseudo-$F = 31.28$; $p = 0.001$) and Bray–Curtis similarity in ANOSIM ($p \leq 0.001$, $R = 0.68$) (Table 3a).

A CCA biplot (Figure 4a) indicated that sampling points were strongly influenced by tree cover and altitude (Table 5, Axis 1). Clear distinctions were found between grassland and savanna sampling points in the CCA ordination, with increased tree cover being correlated with savanna plant assemblages. Forward selection results showed that all tested environmental factors contributed significantly ($p = 0.002$) to variability of the ordination (Table 6). However, the effect of tree cover (a differentiating factor between grassland and savanna) was reduced with the inclusion of longitude, latitude and altitude variables (Table 6).

Plant assemblages across topographic regions

The NMDS results showed that sampling points for plants from each topographic region clustered together to some extent (Figure 3a). PERMANOVA confirmed the distinctions between topographic regions (pseudo-$F = 14.39$; $p = 0.001$). PERMANOVA also revealed a significant interaction between biome and topographic region (pseudo-$F = 14.83$; $p = 0.001$). Samples of savanna localities were more dispersed between sub-escarpment, plateau and escarpment grassland, in terms of plant species composition (Figure 3a and Figure 4). This was confirmed by ANOSIM analyses, which indicated higher *R*-values (greater distinctiveness) for comparisons between savanna localities than between grassland localities (Table 3a). Grassland sampling points revealed limited distinctiveness of plant species composition between topographic regions (Table 3a).

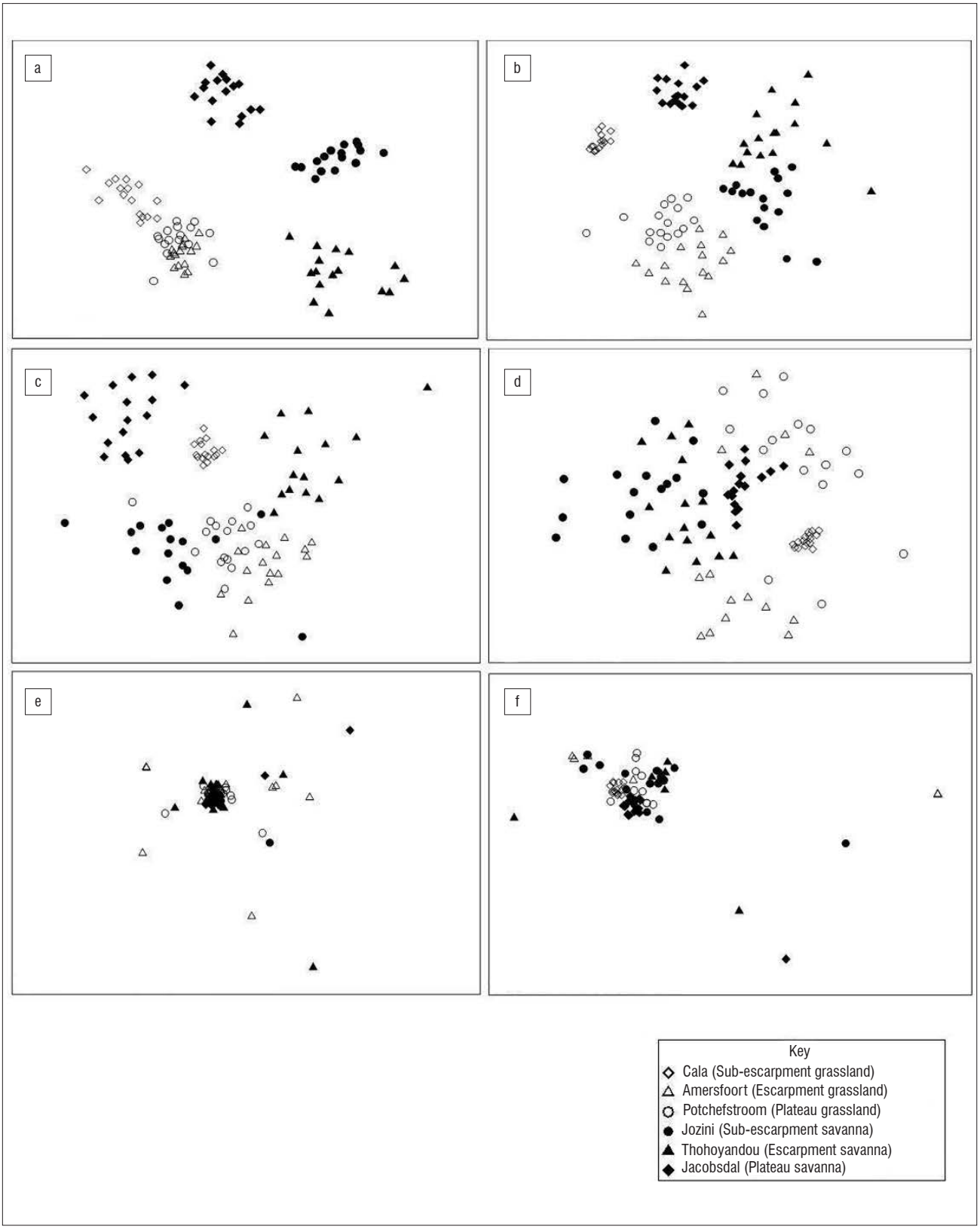


Figure 3: Non-metric multidimensional scaling (NMDS) analyses based on abundance data of plant and arthropod species recorded at maize field margin localities for (a) all plant species, 2D Stress: 0.14, (b) all arthropod species, 2D Stress: 0.23, (c) herbivorous arthropods, 2D Stress: 0.17, (d) predatory arthropods, 2D Stress: 0.16, (e) parasitoid arthropods, 2D Stress: 0.01, and (f) pollinators, 2D Stress: 0.01. Resemblance: S17 Bray–Curtis similarity; data transformation: square root.

Table 3: Results for ANOSIM and SIMPER analyses for biomes and topographic regions in terms of species composition

			ANOSIM *p*-value	ANOSIM *R*-value	SIMPER overall average dissimilarity
a) All plants	Between-biome comparison	Grassland x Savanna	0.0001*	0.68†	97.86
	Within-biome comparison (Grassland)	Escarpment x Plateau	0.0001*	0.52	77.75
		Escarpment x Sub-escarpment	0.0001*	0.94††	91.37
		Plateau x Sub-escarpment	0.0001*	0.81††	88.95
	Within-biome comparison (Savanna)	Escarpment x Plateau	0.0001*	0.99††	99.08
		Escarpment x Sub-escarpment	0.0001*	0.87††	94.25
		Plateau x Sub-escarpment	0.0001*	0.99††	97.23
b) All arthropods	Between-biome comparison	Grassland x Savanna	0.0001*	0.4	98.12
	Within-biome comparison (Grassland)	Escarpment x Plateau	0.0001*	0.5	92.62
		Escarpment x Sub-escarpment	0.0001*	0.98††	99.69
		Plateau x Sub-escarpment	0.0001*	1††	99.41
	Within-biome comparison (Savanna)	Escarpment x Plateau	0.0001*	0.9††	97.8
		Escarpment x Sub-escarpment	0.0001*	0.61†	96.07
		Plateau x Sub-escarpment	0.0001*	0.9††	98.15
c) Herbivores	Between-biome comparison	Grassland x Savanna	0.0001*	0.24	97.94
	Within-biome comparison (Grassland)	Escarpment x Plateau	0.0001*	0.39	90.33
		Escarpment x Sub-escarpment	0.0001*	0.94††	99.35
		Plateau x Sub-escarpment	0.0001*	0.97††	98.88
	Within-biome comparison (Savanna)	Escarpment x Plateau	0.0001*	0. 82††	99.56
		Escarpment x Sub-escarpment	0.0001*	0.57	98.34
		Plateau x Sub-escarpment	0.0001*	0.83††	99.45
d) Predators	Between-biome comparison	Grassland x Savanna	0.0001*	0.33	97.55
	Within-biome comparison (Grassland)	Escarpment x Plateau	0.0001*	0.28	96.13
		Escarpment x Sub-escarpment	0.0001*	0.49	99.47
		Plateau x Sub-escarpment	0.0001*	0.70††	98.98
	Within-biome comparison (Savanna)	Escarpment x Plateau	0.0001*	0.49	93.25
		Escarpment x Sub-escarpment	0.0001*	0.24	90.1
		Plateau x Sub-escarpment	0.0001*	0.63†	93.48
e) Parasitoids	Between-biome comparison	Grassland x Savanna	0.0001*	0.09	98.76
	Within-biome comparison (Grassland)	Escarpment x Plateau	0.0506	0.05	96.02
		Escarpment x Sub-escarpment	0.0001*	0.56†	99.99
		Plateau x Sub-escarpment	0.0001*	0.56†	99.98
	Within-biome comparison (Savanna)	Escarpment x Plateau	0.0001*	0.33	99.22
		Escarpment x Sub-escarpment	0.0025*	0.12	97.74
		Plateau x Sub-escarpment	0.0001*	0.35	99.76
f) Pollinators	Between-biome comparison	Grassland x Savanna	0.0001*	0.07	94.45
	Within-biome comparison (Grassland)	Escarpment x Plateau	0.0038*	0.11	92.38
		Escarpment x Sub-escarpment	0.0001*	0.53†	96.48
		Plateau x Sub-escarpment	0.0001*	0.44	99.09
	Within-biome comparison (Savanna)	Escarpment x Plateau	0.0001*	0.43	100
		Escarpment x Sub-escarpment	0.0074*	0.12	93.51
		Plateau x Sub-escarpment	0.0001*	0.35	99.54

** significant at $p<0.05$.*

†† large effect at $R\geq0.7$; † medium effect at $R\geq0.5$.

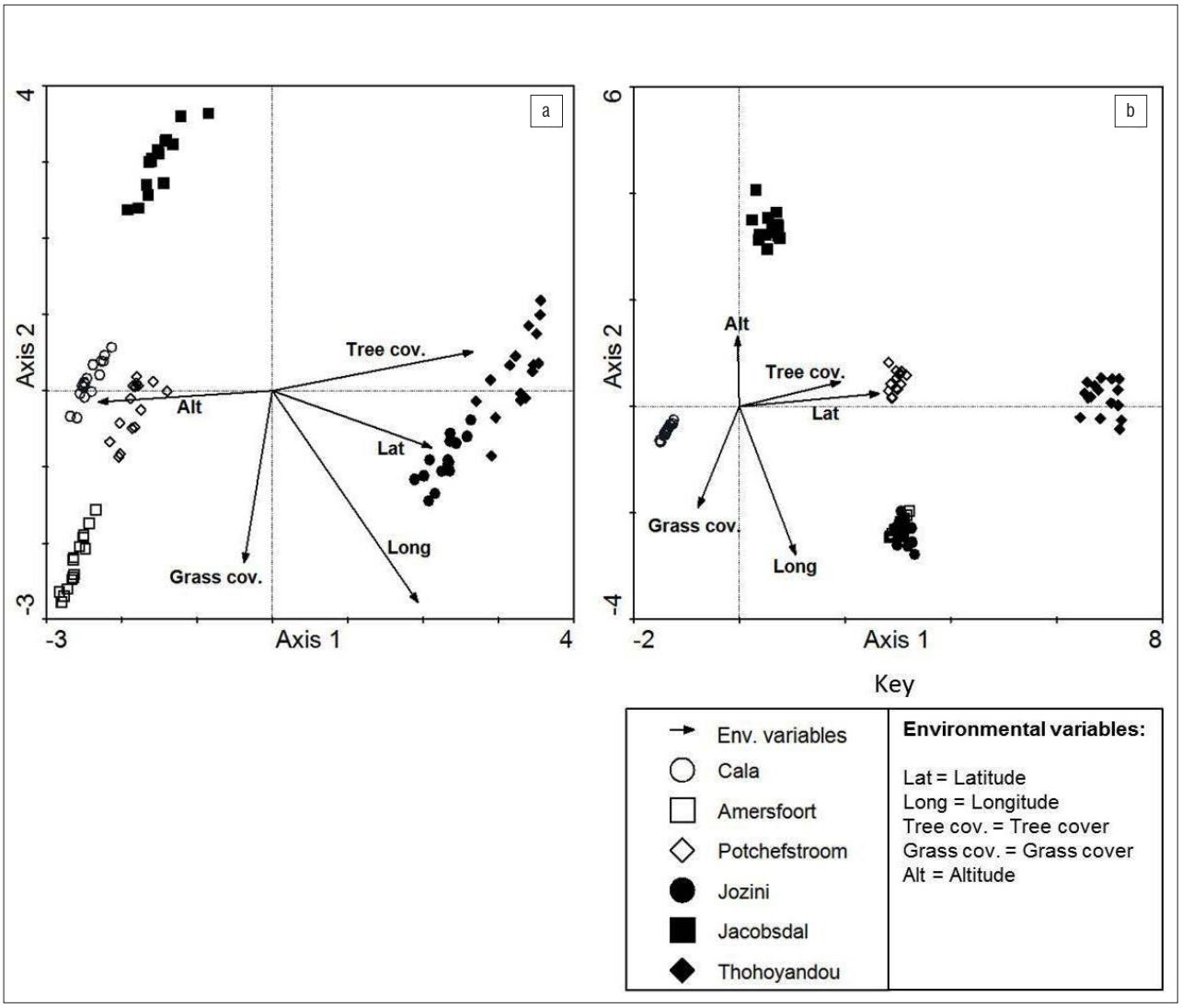


Figure 4: Canonical correspondence ordination of all sampled localities, showing correlations between environmental variables and sampling points for (a) plants and (b) arthropods. Each symbol represents the weighted average of one plot.

Contributing plant species

A total of 21 plant species each contributed more than 1% to the total variation between the biomes, according to SIMPER analysis. Among these, ten species collectively contributed 21% of the variation (Table 4). Six grasses (*Eragrostis curvula, Eragrostis plana, Heteropogon contortus, Hyparrhenia hirta, Sporobolus africanus* and *Themeda triandra*) and one forb (*Helichrysum rugulosum)* characterised the grassland points, whereas one forb (*Pentzia incana*) and two grasses (*Panicum maximum* and *Urochloa mosambicensis*) characterised the savanna points (Table 4).

Arthropod assemblages across biomes

In the NMDS analysis for all arthropod species, grassland sampling points were noted to cluster separately from savanna points (Figure 3b). The biomes, however, were more similar in arthropod species composition than plant species composition, as was evident from the PERMANOVA results (arthropods: pseudo-$F=9.712$; $p=0.001$; plants: pseudo-$F=31.278$, $p=0.001$). ANOSIM showed similar results, indicating less similarity among arthropods ($p \leq 0.001$; $R=0.4$) (Table 3b).

The CCA results for arthropods (Figure 4b) showed that sampling points were more strongly influenced by latitude (i.e. geographic position) than either plant cover or altitude (Axis 1, Table 5). Distinctions were not as clear between grassland and savanna sampling points for arthropods as they were for plants. Forward selection results confirmed the CCA ordination results, which suggests that longitude and latitude explained the most variation when all environmental factors were considered ($p=0.002$) (Table 6). This finding shows a notable division between samples taken from the eastern and western regions of South Africa.

Arthropod assemblages across topographic regions

For the complete arthropod dataset, sampling points from the same topographic region tended to cluster together (Figure 3b). This pattern was also evident in PERMANOVA results, which indicated significant differences between the regions (pseudo-$F=8.009$; $p=0.001$). PERMANOVA showed a smaller interactive effect between biome and topographic region for arthropods than for plants (arthropods: pseudo-$F=8.823$, $p=0.001$; plants: pseudo-$F=14.834$, $p=0.001$). Sampling points from sub-escarpment and escarpment localities in the eastern half of South Africa clustered more closely for savanna than for grassland. However, plateau and escarpment points in the eastern region of South Africa clustered together more strongly for grassland.

Contributing arthropod morpho-species

In the ordination of the total arthropod dataset, a total of ten arthropod morpho-species each contributed more than 1% to the total variation across the biomes, according to SIMPER analysis (Table 4).

Table 4: Results for SIMPER analyses indicating the top ten plant and arthropod species responsible for groupings of grassland and savanna plots in the NMDS graphs

	Species	Type	Ave. dis.	% Contr.	Cumul. %	Abund. Grassland	Abund. Savanna
Plants (overall average dissimilarity: 97.86)	*Themeda triandra*	Grass	3.684	3.764	3.764	8.27	1.6
	Panicum maximum	Grass	3.246	3.317	7.082	0	7.6
	Pentzia incana	Forb	2.355	2.406	9.488	0	4.88
	Helichrysum rugulosum	Forb	1.977	2.02	11.51	4.5	0
	Eragrostis curvula	Grass	1.753	1.791	13.3	3.25	1.27
	Eragrostis plana	Grass	1.741	1.779	15.08	3.94	0
	Sporobolus africanus	Grass	1.704	1.741	16.82	3.79	0.292
	Heteropogon contortus	Grass	1.45	1.482	18.3	2.83	0.979
	Hyparrhenia hirta	Grass	1.444	1.475	19.78	3.25	0
	Urochloa mosambicensis	Grass	1.438	1.47	21.25	0	3.21
Arthropods (overall average dissimilarity: 98.12)	Oribatulidae MS1	Detritivore	3.327	3.391	3.391	63	0.125
	Oribatulidae MS2	Detritivore	3.081	3.14	6.531	66.8	0
	Formicidae MS8	Predator	2.675	2.726	9.257	2.63	2.17
	Acrididae MS12	Herbivore	2.384	2.43	11.69	2.25	0.896
	Cicadellidae MS8	Herbivore	2.266	2.309	14	6.6	0.458
	Entomobryidae MS1	Detritivore	2.081	2.121	16.12	35	0.146
	Cicadellidae MS24	Herbivore	1.367	1.393	17.51	0	2.65
	Sciaridae MS10	Detritivore	1.233	1.257	18.77	0	4.79
	Cecidomyiidae MS1	Herbivore	1.227	1.251	20.02	17	1.02
	Formicidae MS9	Predator	1.17	1.192	21.21	14.8	0.313

Key to column headings: Ave. dis, average dissimilarity; % Contr., percentage contribution of each species to the average dissimilarity; Cumul. %, cumulative contribution percentage; Abund., mean abundance per plot.

Table 5: Correlations of ordination axes with selected environmental factors, eigenvalues and percentage variance explained for canonical correspondence analysis

Survey	Factor	Axis 1	Axis 2
Plants	Tree cover %	0.8975	0.1274
	Altitude	-0.7770	-0.0368
	Latitude	0.7133	0.1854
	Longitude	0.6511	-0.6904
	Grass cover %	-0.1252	-0.5610
	Eigenvalue	0.860	0.702
	% variance explained	29.8	24.3
Arthropods	Latitude	0.9866	0.0762
	Tree cover %	0.7128	0.1462
	Longitude	0.3969	-0.8754
	Grass cover %	-0.2880	-0.5962
	Altitude	-0.0126	0.4150
	Eigenvalue	0.824	0.717
	% variance explained	28.8	25.1

Collectively these species contributed 21% of all variation across biomes. Two mite (Oribatulidae), one springtail (Entomobryoidea), one gall gnat (Cecidomyiidae) and one ant (Formicidae) species characterised the grassland sampling points, whereas one leafhopper (Cicadellidae) and one fungus gnat (Sciaridae) species supported the distinctiveness of the savanna sampling points (Table 4).

Assemblages of arthropod trophic groups

The NMDS analyses for separate arthropod trophic groups showed a much more uniform distribution of sampling points than that of plants or the complete arthropod dataset (Figures 3c to f). However, PERMANOVA showed that differences in species composition between biomes were still significant for all arthropod trophic groups (herbivores: pseudo-F=9.013, p=0.001; predators: pseudo-F=12.317, p=0.001; parasitoids: pseudo-F=5.566, p=0.001; pollinators: pseudo-F=3.507, p=0.001). These results also show that the predators and herbivores displayed the most significant differences in species composition between biomes. The results were confirmed by the R-values of ANOSIM analyses (Table 3c to f).

No clearly distinctive clustering could be observed for topographic regions for any of the arthropod trophic groups in the NMDS analyses. However, PERMANOVA indicated that there were significant differences in species composition between some topographic regions for all the trophic groups (herbivores: pseudo-F=8.127, p=0.001; predators: pseudo-F=8.207, p=0.001; parasitoids: pseudo-F=5.126, p=0.001; pollinators: pseudo-F=3.882, p=0.001).

Table 6: Marginal and conditional effects of automatic forward selection conducted for all plants and all arthropods

	Marginal effects			Conditional effects				
	Variable	**Var. N**	**Lambda1**	**Variable**	**Var. N**	**LambdaA**	***P***	***F***
Plants	Altitude	1	0.79	Altitude	1	0.79	0.002*	4.51
	Tree cover	2	0.77	Latitude	3	0.74	0.002*	4.42
	Latitude	3	0.77	Longitude	4	0.66	0.002*	4.06
	Longitude	4	0.76	Grass cover	5	0.39	0.002*	2.42
	Grass cover	5	0.49	Tree cover	2	0.31	0.002*	1.94
Arthropods	Latitude	1	0.82	Latitude	1	0.82	0.002*	2.71
	Longitude	2	0.73	Longitude	2	0.71	0.002*	2.37
	Tree cover	3	0.68	Altitude	4	0.63	0.002*	2.15
	Altitude	4	0.65	Tree cover	3	0.39	0.004*	1.32
	Grass cover	5	0.50	Grass cover	5	0.31	0.364	1.03

* significant p-values ($p<0.05$) as determined by Monte Carlo permutation tests (permutations=499)

These results show that the largest distinctions in species composition between topographic regions for the biomes were within the herbivore and predator groups (in both cases mainly between the plateau and sub-escarpment). For parasitoids and pollinators, there was almost complete species homogeneity between regions, as confirmed by ANOSIM ($R \leq 0.09$) (Table 3e and f). Across biomes, escarpment localities showed some clustering of points for herbivores (Figure 3c). Escarpment and plateau respectively showed clusters for predators across the two biomes (Figure 3d). Sub-escarpment localities had the most dissimilar arthropod composition between grassland (Cala) and savanna (Jozini) biomes, for herbivores and predators.

PERMANOVA also indicated significant interaction effects between biome and topographic region for all trophic groups (herbivores: pseudo-$F=9.463$, $p=0.001$; predators: pseudo-$F=7.462$, $p=0.001$; parasitoids: pseudo-$F=5.85$, $p=0.001$; pollinators: pseudo-$F=4.093$, $p=0.001$). This pattern was again more evident for herbivores and predators than for parasitoids and pollinators.

Discussion

Our results confirmed findings that between biomes, plant species assemblages were more distinguishable than arthropod assemblages.[11] The ordination of plant species had the tightest clustering and lowest stress value (0.14). The NMDS analysis for all arthropod morpho-species had a high stress value (0.23), which indicates that not too much reliance can be placed on the spacing of plots of the ordination without cross-checking the results through other statistical analyses.[25] However, the similar patterns in species composition between this ordination and other NMDS analyses (Figures 3c to f), which yielded low stress values, suggest that this plot is a reasonably good representation of the relationships between samples, and can be further interpreted.

Arthropod assemblages, based on all morpho-species, seem to cluster according to biomes at least to some extent, which is consistent with previous results[11]. The effect of tree cover (a differentiating factor between grassland and savanna) remained noteworthy even with the inclusion of longitude, latitude and altitude variables in the CCA forward selection (Table 6). Possible causes of the separate clusters of grassland and savanna arthropod assemblages can be ascribed to preferences for specific climatic conditions associated with the biomes.[29] However, the distinction between grassland and savanna arthropod species assemblages was far less marked than it was for plants. This distinction was even smaller for separate trophic groups.

The phytophagous and predacious arthropod groups showed the most distinctive groupings between the biomes, of all the trophic groups. This finding suggests that these groups are more specialised and adapted to the two biomes than are parasitoids, pollinators or other trophic groups.[14,29,30] High levels of similarity between the biomes with regard to insect species assemblages for parasitoids and pollinators can be ascribed to the extreme mobility of this group; many of these species are capable of flight[11]. In our study, the homogenisation effect for arthropods in these groups is probably related more to plant phylogenies of the biomes and host plant specificity.[3] These host plant taxa could be typical of either savanna or grassland, which could lead to homogenisation of the insect groups.

Our results seem to indicate that arthropod assemblages are better explained by their geographical position, particularly longitude and latitude, than by biome characteristics. This finding could be an effect of altitude (localities generally increase in altitude from north to south) and climate (localities become drier from east to west). In line with this reasoning, relatively distinct plant and arthropod assemblages occurred within each topographic region. The effect of altitude on species composition is well known.[31] Arthropod communities at different elevations often experience markedly different environmental conditions, particularly climatic. Several studies have demonstrated the dependence of species composition on altitude for several arthropod groups.[32-34]

Conclusion

When all arthropods are considered, the grassland and savanna biomes have distinct arthropod assemblages. However, the degree of dissimilarity among plant assemblages is greater between grassland and savanna biomes. When trophic levels were compared, the distinction between arthropod assemblages became even more obscured between biomes. Biomes were still distinguishable, albeit weakly, for phytophagous and predacious arthropod assemblages, but not for parasitoids and pollinators. The similarity in arthropod assemblages for different trophic levels can be ascribed to both biomes being characterised by a dominant grass layer and hence habitats. Arthropod species assemblages were better explained by their geographical position than by plant features associated with biome, such as tree and grass cover. It must be noted that our results are based on a limited range of environmental factors and species groups, and further research is required to confirm these patterns for arthropods under different conditions and spatial scales.

Authors' contributions

S.J.S. and J.v.d.B. were the project leaders, and were responsible for experimental and project design. All authors contributed towards data collection. M.B. and S.J.S. performed the statistical analyses and wrote the manuscript.

Acknowledgements

This work formed part of the Environmental Biosafety Cooperation Project between South Africa and Norway, coordinated by the South African National Biodiversity Institute. Financial support was also provided by GenØk-Centre of Biosafety, Norway, Norad project GLO-3450. North-West University in South Africa provided logistical support. B.G. Maliba is thanked for making his published data available and for initial discussions on the topic.

References

1. Rutherford MC, Westfall RH. Biomes of southern Africa: An objective characterisation. Mem Bot Surv S Afr. 1986;54:1–98.
2. Mucina L, Rutherford MC, editors. The vegetation of South Africa, Lesotho and Swaziland. Strelitzia 19. Pretoria: South African National Biodiversity Institute; 2006. [Reprint 2011; future update to occur on 9 October 2016].
3. Moolman J, van den Berg J, Conlong D, Cugala D, Siebert S, Le Ru B. Species diversity and distribution of lepidopteran stem borers in South Africa and Mozambique. J Appl Entomol. 2013;137(1–2):641–720. http://dx.doi.org/10.1111/jen.12085
4. Schaffers AP, Raemakers IP, Sykora KV, Braak CJFT. Arthropod assemblages are best predicted by plant species composition. Ecology. 2008;89(3):782–794. http://dx.doi.org/10.1890/07-0361.1
5. Symstad AJ, Siemann E, Haarstad J. An experimental test of the effect of plant functional group diversity on arthropod diversity. Oikos. 2000;89(2):243–253. http://dx.doi.org/10.1034/j.1600-0706.2000.890204.x
6. Batáry P, Orci KM, Báldi A, Kleijn D, Kisbenedek T, Erdős S. Effects of local and landscape scale and cattle grazing intensity on Orthoptera assemblages of the Hungarian Great Plain. Basic Appl Ecol. 2007;8(3):280–290. http://dx.doi.org/10.1016/j.baae.2006.03.012
7. Gebeyehu S, Samways M. Grasshopper assemblage response to a restored national park (Mountain Zebra National Park, South Africa). Biodivers Conserv. 2002;11(2):283–304. http://dx.doi.org/10.1023/A:1014514007744
8. Kuppler J, Fricke J, Hemp C, Steffan-Dewenter I, Peters MK. Conversion of savannah habitats to small-scale agriculture affects grasshopper communities at Mt. Kilimanjaro, Tanzania. J Insect Conserv. 2015;19(3):509–518. http://dx.doi.org/10.1007/s10841-015-9772-7
9. Pryke JS, Roets F, Samways MJ. Importance of habitat heterogeneity in remnant patches for conserving dung beetles. Biodivers Conserv. 2013;22(12):2857–2873 http://dx.doi.org/10.1007/s10531-013-0559-4
10. Proches S, Cowling RM. Insect diversity in Cape fynbos and neighbouring South African vegetation. Global Ecol Biogeog. 2006;15(5):445–451. http://dx.doi.org/10.1111/j.1466-822X.2006.00239.x
11. Proches S, Cowling RM. Do insect distributions fit our biomes? S Afr J Sci. 2007;103(5/6):258–261.
12. Wright MG, Samways MJ. Insect species richness tracking plant species richness in a diverse flora: Gall-insects in the Cape Floristic Region, South Africa. Oecologia 1998;115(3):427–433. http://dx.doi.org/10.1007/s004420050537
13. Botha M, Siebert SJ, van den Berg J, Maliba BG, Ellis SM. Plant and arthropod diversity patterns of maize agro-ecosystems in two grassy biomes of South Africa. Biodivers Conserv. 2015;24(7):1797–1824. http://dx.doi.org/10.1007/s10531-015-0901-0
14. Krüger O, McGavin G. The influence of ants on the guild structure of *Acacia* insect communities in Mkomazi game reserve, north-east Tanzania. Afr J Ecol. 1998;36(3):213–220. http://dx.doi.org/10.1046/j.1365-2028.1998.00130.x
15. Schadek U, Strauss B, Biedermann R, Kleyer M. Plant species richness, vegetation structure and soil resources of urban brownfield sites linked to successional age. Urban Ecosyst. 2009;12(2):115–126. http://dx.doi.org/10.1007/s11252-008-0072-9
16. Koricheva J, Mulder CP, Schmid B, Joshi J, Huss-Danell K. Numerical responses of different trophic groups of invertebrates to manipulations of plant diversity in grasslands. Oecologia. 2000;125(2):271–282. http://dx.doi.org/10.1007/s004420000450
17. Ricklefs RE, Marquis RJ. Species richness and niche space for temperate and tropical folivores. Oecologia. 2012;168(1):213–220. http://dx.doi.org/10.1007/s00442-011-2079-9
18. Tews J, Brose U, Grimm V, Tielborger K, Wichmann MC, Schwager M, et al. Animal species diversity driven by habitat heterogeneity/diversity: The importance of keystone structures. J Biogeogr 2004;31(1):79–92. http://dx.doi.org/10.1046/j.0305-0270.2003.00994.x
19. Elliott NC, Tao FL, Fuentes-Granados R, Giles KL, Elliot DT, Greenstone MH, et al. D-vac sampling for predatory arthropods in winter wheat. Biol Control 2006;38(3):325–330. http://dx.doi.org/10.1016/j.biocontrol.2006.04.017
20. Thomas CFG, Marshall EJP. Arthropod abundance and diversity in differently vegetated margins of arable fields. Agric Ecosyst Environ. 1999;72(2):131–144. http://dx.doi.org/10.1016/S0167-8809(98)00169-8
21. Hill DA. Handbook of biodiversity methods: Survey, evaluation and monitoring. Cambridge: Cambridge University Press; 2005. http://dx.doi.org/10.1017/CBO9780511542084
22. Ugland KI, Gray JS, Ellingsen KE. The species–accumulation curve and estimation of species richness. J Anim Ecol. 2003;72(5):888–897. http://dx.doi.org/10.1046/j.1365-2656.2003.00748.x
23. Hortal J, Borges PA, Gaspar C. Evaluating the performance of species richness estimators: Sensitivity to sample grain size. J Anim Ecol. 2006;75(1):274–287 http://dx.doi.org/10.1111/j.1365-2656.2006.01048.x
24. Clarke KR, Gorley RN. PRIMER v6: User manual / tutorial. [Article and software on the Internet.] Available from http://www.primer-e.com. Plymouth: PRIMER-E; 2006.
25. Clarke KR, Warwick RM. Change in marine communities: An approach to statistical analysis and interpretation. 2nd ed. Plymouth: PRIMER-E; 2001.
26. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001;26(1):32–46. http://dx.doi.org/10.1111/j.1442-9993.2001.01070.pp.x
27. Clarke KR. Non-parametric multivariate analyses of changes in community structure. Aust J Ecol. 1993;18(1):117–143. http://dx.doi.org/10.1111/j.1442-9993.1993.tb00438.x
28. Ter Braak CJ, Verdonschot PF. Canonical correspondence analysis and related multivariate methods in aquatic ecology. Aquat Sci. 1995;57(3):255–289. http://dx.doi.org/10.1007/BF00877430
29. Menéndez R, González-Megías A, Collingham Y, Fox R, Roy DB, Ohlemüller R, et al. Direct and indirect effects of climate and habitat factors on butterfly diversity. Ecology 2007;88(3):605–611. http://dx.doi.org/10.1890/06-0539
30. Mitter C, Farrell BD, Wiegmann B. The phylogenetic study of adaptive zones: Has phytophagy promoted insect diversification? Am Nat. 1988;132(1):107–128. http://dx.doi.org/10.1086/284840
31. Hodkinson ID. Terrestrial insects along elevation gradients: Species and community responses to altitude. Biol Rev Camb Philos Soc. 2005;80(3):489–513. http://dx.doi.org/10.1017/S1464793105006767
32. Adams M, Bernard G. Pronophiline butterflies (Satyridae) of the Cordillera de Mérida, Venezuela. Zool J Linn Soc. 1981;71(4):343–372. http://dx.doi.org/10.1111/j.1096-3642.1981.tb01135.x
33. Eyre M, Woodward J, Luff M. The distribution of grassland Auchenorrhyncha assemblages (Homoptera: Cercopidae, Cicadellidae, Delphacidae) in northern England and Scotland. J Insect Conserv. 2001;5(1):37–45. http://dx.doi.org/10.1023/A:1011330624531
34. Gutiérrez D, Menéndez R. Distribution and abundance of butterflies in a mountain area in the northern Iberian peninsula. Ecography 1995;18(3):209–216. http://dx.doi.org/10.1111/j.1600-0587.1995.tb00123.x

Assessment of visualisation skills in biochemistry students

AUTHORS:
Lindelani Mnguni[1]
Konrad Schönborn[2]
Trevor Anderson[3]

AFFILIATIONS:
[1]Department of Science and Technology Education, University of Johannesburg, Johannesburg, South Africa
[2]Visual Information Technology and Applications (VITA), Department of Science and Technology, Linköping University, Linköping, Sweden
[3]Visualization in Biochemistry Education (VIBE) Research Group, Department of Chemistry, University of Purdue, West Lafayette, Indiana, United States

CORRESPONDENCE TO:
Lindelani Mnguni

EMAIL:
lmnguni@uj.ac.za

DATES:
Received: 04 Nov. 2015
Revised: 15 Feb. 2016
Accepted: 12 Apr. 2016

KEYWORDS:
external representations; Rasch model; visual literacy

HOW TO CITE:
Mnguni L, Schönborn K, Anderson T. Assessment of visualisation skills in biochemistry students. S Afr J Sci. 2016;112(9/10), Art. #2015-0412, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20150412

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

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

In the field of biochemistry, the use of external representations such as static diagrams and animations has increased rapidly in recent years. However, their effectiveness as instructional tools can be hindered if students lack the visual literacy and cognitive skills necessary for processing and interpreting such representations. We aimed to identify and assess visualisation skills necessary for effective processing of external representations in biochemistry. We used a modified Bloom's taxonomy to identify the cognitive skills essential for optimal visual literacy, and designed probes based on those skills to develop a test instrument. Student responses to the probes were scored and processed with the Rasch model. This approach enabled us to rate the degree of difficulty of each visualisation skill on a linear logit scale, and to generate a person–item map to measure biochemistry students' level of visual literacy. The results showed that the identified visualisation skills could be measured reliably, and the Rasch model was effective both for ranking the skills according to level of difficulty and for estimating a student's relative level of visual literacy.

Significance:

- Addresses a recurring problem in biochemistry and similar fields.
- Identifies relevant skills to inform teaching and learning in biochemistry.

Introduction

Bloom's taxonomy is widely accepted as the golden standard for determining learning objectives. In its purest form, Bloom's taxonomy addresses cognitive, affective and psychomotor domains, and can be applied to various competencies. Competencies include mathematics literacy, science literacy and visual literacy (VL), with VL being a key competency in learning biochemistry. According to Mnguni[1], the visualisation process has three main stages: internalisation, conceptualisation and externalisation. In this model, internalisation refers to the process in which sense organs (such as the eyes) work with the brain to absorb information from the world (i.e. external to the body). Conceptualisation is the process by which meaning is made and cognitive visual models are constructed.[2] During conceptualisation, prior knowledge that has been stored as cognitive visual models may be retrieved from long-term memory and reconstructed or revised in working memory, based on new knowledge. Externalisation is the production of external visual models by way of expressing cognitive mental schema.

Learning in biochemistry is complicated by a number of factors. For example, many biomolecular phenomena cannot be visualised with the naked eye because of their submicroscopic size and levels of complexity. Furthermore, these often abstract phenomena occur across various levels of organisation, usually at molecular level.[3] Students therefore have to learn these concepts through individualised cognitive negotiation of 'imagined' concepts. Teaching, in such a field, requires perfecting the art of imagination amongst students and promoting uniformity in the cognitive images formed through the use of formalised external representations (ERs). To this end, a variety of ERs – such as static diagrams and animations – are used to express phenomena graphically and to assist students in visualising phenomena and constructing knowledge of how those phenomena occur in reality.[4] However, students often fail to correctly visualise and interpret ERs in a manner that provides them with sound conceptual understanding.[5,6] It is therefore not surprising that a lack of VL is one of the major problems faced by students studying biochemistry in modern educational settings.[7]

Schonbörn and Anderson[3] argue that students find it challenging to master the abstract and diverse symbolic language used to represent and externalise biomolecular phenomena. This problem is compounded by the fact that some teaching and learning tools which experts consider to be good are not always effective at promoting learning in novices.[8] In addition, students lack the cognitive skills required for optimal (expert level) VL skills which are needed to process and construct meaning from ERs.[9,10] In line with this argument, some educators feel that students do not need to be explicitly 'taught' the visualisation skills necessary for interpreting ERs, and instead assume that such skills are automatically acquired during the imparting of prescribed content knowledge.[11] However, several studies have suggested this is certainly not always true.[12,13] We therefore sought to investigate and assess the VL of students studying biochemistry, so that meaningful action can be taken to help students process ERs effectively[3].

Purpose of the study

Our research question was 'What visualisation skills are required for students to effectively process ERs related to biochemistry content?' We hypothesised that a typical class of biochemistry students would show a range of VL levels, which would depend on various factors and could improve with gains in conceptual knowledge and competence in visualisation skills. Given this hypothesis, the specific aims of the study were: (1) to establish which cognitive skills are important for VL in biochemists; (2) to develop a test instrument to assess each relevant visualisation skill; (3) to process student responses with the Rasch model and rank visualisation skills in order of difficulty; and (4) to use the Rasch model to construct a person–item map to determine students' relative levels of VL. To address these aims, we

employed a primarily quantitative design. Elements of a mixed-methods approach were also adopted, in which both qualitative and quantitative data were collected and analysed.

Our main objective was to develop an instrument to test students' VL in biochemistry, based on item–response theory using the Rasch model. We used item–response theory because it allows for a number of features that are not provided by other forms, such as classical test theory. One such feature is that item–response theory converts non-linear raw scores to linear logit values that can be used to control for the difficulty level of scale items and the non-additive feature of ordinal data.[14,15] In this way, 'responses based on the ordinal items are transformed into an interval scale based on logits to which proper parametric statistics can be applied'.[14] Logits are interpreted directly; for instance, an item with 20 logits is twice as difficult as an item with 10 logits. A person with 10 logits has double the ability of a person with 5 logits. This comparison is impossible with normal classical test theory, in which a person who scores 50% does not necessarily have twice the ability as a person who scores 25%.

The Rasch model

To generate logit scores, a number of item–response theory models can be used. One common model is the single parameter Rasch model, which uses mathematical formulae to calculate probabilities based on the actual scores obtained.[15] These probability-based scores are then converted to logit scores. Using fit statistics, the model is able to detect logit scores that differ too widely from actual scores before conversion. In the Rasch model, fit statistics indicate how accurately data fit the model. In this instance, 'infit' means inlier-sensitive and 'outfit' means outlier-sensitive fit, whereas mean-square fit statistics show the degree of randomness. Most scholars[16] suggest that ideal fit statistics should be around 1.5. Extreme differences from this value indicate a less reliable instrument.

Another feature of the Rasch model is that it can determine dimensionality of the test. In this regard, tests measuring psychological variables (such as ability) on an interval scale should measure exactly the same variable with equal intervals in the level of difficulty. In other words, other factors that may tamper with that variable should be eliminated. The test should be unidimensional – that is, it should measure one variable only. The Rasch model determines dimensionality by calculating the Rasch unidimensionality coefficient. Smith et al.[17] argue that absolute unidimensionality is observed if the mean-square fit statistic equals 1. They state that if the fit statistic of an item is 1.25, this indicates 25% variation; if the fit statistic is 0.7 then there is 30% less variation. However, other scholars[16,18] suggest that fit statistics ranging from 0.5 to 1.7 are acceptable indicators of unidimensionality in clinical observations.

Cognitive skills that contribute to visual literacy

We consulted the literature to identify cognitive skills that are intrinsic to VL. According to Mnguni[1], the main stages of the visualisation process are internalisation, conceptualisation and externalisation. Based on this model, we used the revised Bloom's taxonomy[19] to identify cognitive skills that would most likely be engaged during each stage. We chose the revised Bloom's taxonomy because of its wide application as a mechanism of classification and categorisation of levels of learning.[19]

We studied each cognitive skill from Bloom's taxonomy with respect to examples of activities carried out and the definition of each cognitive skill. Thereafter we placed each cognitive skill in a visualisation stage that is most relevant. As shown in Table 1, this procedure yielded 24 cognitive skills, which we called *visualisation skills* because of their association with the visualisation process. As shown in Table 1, each skill has its own code number. Importantly, we realised that not all skills are utilised exclusively during a single stage of the visualisation process. For instance, 'mental rotation' (code T16) requires first perceiving the orientation of an object, followed by cognitive processes to mentally rotate the object in different dimensions.

Development of instrument to assess visualisation skills

We used the visualisation skills identified (Table 1) to develop an instrument to assess students' visualisation skills. We noted that students' interpretation of ERs in the test instrument would depend on at least three *interdependent* factors. These are students' reasoning ability (including visualisation skills) to make sense of an ER, prior conceptual knowledge that students apply in interpreting the ER, and the mode in which the knowledge is represented externally (e.g. graphic features, markings, diagrams and animations).[20] The interdependent nature of these factors during the visualisation process suggests that it is not possible to design an assessment task that would assess students' visualisation skills exclusively, without also considering students' conceptual knowledge of relevance to a particular ER.

We therefore developed ER-based probes that require respondents to utilise their reasoning skills and prior knowledge in their search for answers. For each visualisation skill we first identified a basic biochemistry concept that in our judgement requires the use of a particular visualisation skill. Thereafter we identified an ER that is associated with each biochemistry concept. We then developed a probe that requires utilisation of the visualisation skill, by testing students' understanding of the biochemistry concept as represented by an ER. Each probe focused on at least one aspect of basic biochemistry, which included the topics of amino acid and protein structure, nucleic acid and protein synthesis, cellular structures and protein binding.

To minimise the influence of conceptual knowledge on assessing students' visualisation skill competence and therefore any VL measurements, we kept the biochemistry content required to answer the probes as basic as possible. This ensured that participating students would have few problems understanding the content. We also obtained expert validation of the test instrument, as discussed later in this paper. In addition, we examined students' performance on other course tests in which similar propositional knowledge was assessed. This analysis revealed that our student sample had adequate knowledge of the basic biochemistry concepts tested by the probes.

The final test instrument consisted of 12 probes. Each probe assessed more than one of the 24 visualisation skills. All probes included an ER, such as a static diagram or animation, with accompanying text. The probes were presented to students in MS PowerPoint format, with each question having an allotted time in which it had to be answered. For example, in probe 1, students were asked to determine whether two diagrams depicting amino acid features (Figure 1) represented the same amino acid. In this question, students were expected to use visualisation skills and knowledge of symbolic language to explain the concepts represented by the various symbols, graphical markings and visual cues composing the ERs. In essence, the visualisation skill 'analyse' (code T01, Table 1) was the skill of interest, because students were expected to break the ERs down into smaller components or essential features and provide a detailed explanation of each through careful examination.

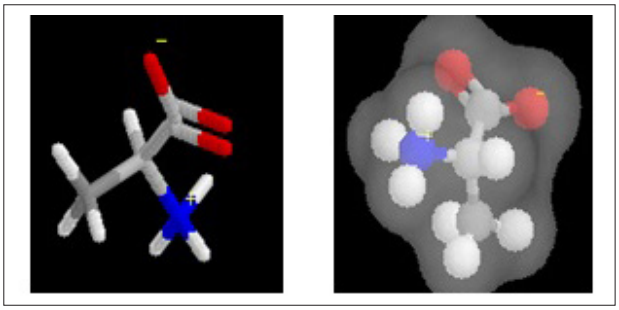


Figure 1: Two different visual representations of the same amino acid used in Probe 1.

Table 1: List of visualisation skills required for visual literacy

Visualisation stage	Visualisation skills	Visualisation skills definition	Visualisation skill code
Internalisation	Arrange; order; organise; classify	To put into a specific order or relation through a methodical or systematic arrangement, or to arrange in a coherent form or pattern based on specific features	T02
	Depth perception; recognition of depth cues	To perceive spatial relationships and distances between objects, in multi-dimensions	T06
	Find; locate	To come upon or discover by searching or making an effort; to discover or ascertain through observation; to determine or specify the position or limits of by searching, examining.	T09
	Focus	To concentrate attention energy on something	T10
	Ground perception	To detect or perceive the part of a scene (or picture) that lies behind objects in the foreground	T11
	Perceive luminance; identify colours	To detect or perceive a visual attribute of things that result from the light they emit or transmit or reflect	T18
	Perceive motion	To recognise, discern, envision, or understand change of position in space and assign meaning to	T19
	Perceive speed	To recognise, discern, envision, or understand a rate of movement and meaning thereof	T20
	Perceive texture	To recognise, discern, envision, or understand the characteristic visual and tactile quality of the surface and meaning of such	T21
Conceptualisation	Analyse; interpret; assess; evaluate; examine; investigate	To break down into components or essential features by making sense of or assigning a meaning to, or give explanation and to examine or assess carefully; observe or inquire into in detail, by examining systematically; to observe carefully or critically.	T01
	Compare; relate	To examine and note the similarities or differences; bring into or link in logical or natural association, and establish or demonstrate a connection between	T03
	Critique	To critically examine and judge something	T05
	Imagine	To form a mental image of something that is not present or that is not given	T13
	Describe; discuss; explain	To make plain or comprehensible by adding details; to justify or offer reasons for or a cause, and give a description of, by conveying an idea or impression in speech or writing; characterise	T07
	Discriminate	To recognise or perceive the difference	T08
	Judge	To determine or declare after consideration or deliberation; to form an opinion or evaluation	T15
	Manipulate; mental rotation; recognise orientation; recognition; identify; identify shapes	To move, arrange, operate, or control cognitively in a skilful manner for examination purposes, and then to perceive multiple items with different orientation and/or shape to be the same if orientation and/or shape is rearranged	T16
	Recall; retrieve	To remember by retrieving information from memory	T23
Externalisation	Complete	To make whole, with all necessary or normal elements or parts	T04
	Illustrate; sketch	To clarify, as by use of examples or comparisons and to use drawings to describe roughly or briefly or give the main points or summary of	T12
	Infer; predict	To conclude by reasoning; in logic or reason or establish by deduction or state, tell about, or make known in advance, on the basis of special knowledge	T14
	Outline	To give the main features or various aspects of; summarise	T17
	Propose; develop; formulate; devise; construct; create; produce; invent	To cause to exist in a new or different form through artistic or imaginative effort	T22
	Use	To put into service or apply for a purpose	T24

For the skill 'analyse' (T01), students' responses were scored as 'correct' (3 points) if there was evidence of visualisation skills and correct interpretation of the ER. Such skills were demonstrated through knowledge of the symbolic language and a high degree of conceptual understanding of the biochemistry (propositional) knowledge represented by the ER. A score of 'acceptable' (2 points) was given if there was evidence of a moderate level of visualisation and interpretation of the ER, as demonstrated by some understanding of the symbolic language and biochemistry (propositional) knowledge represented by the ER. A score of 'partially correct' (1 point) was given if there was evidence of very limited visualisation and interpretation of the ER, as demonstrated by a poor understanding of the symbolic language and biochemistry (propositional) knowledge represented by the ER. A score of 'incorrect' (zero) was given if there was an incorrect or no response to the question, or an incorrect interpretation of the ER based on inappropriate reasoning and incorrect symbolic and conceptual knowledge.

Scoring was performed by the three researchers independently, and the results were then compared. Differences and concerns were negotiated among the researchers until consensus was reached. By the end of this exercise, each student had been given an aggregate score for each visualisation skill, which enabled the calculation of an overall average score for the entire student sample and for each visualisation skill (Table 1). This approach led to the output of non-linear raw scores, which were then further processed using the WINSTEPS Rasch model, which is a rating scale model often employed to analyse Likert-type data.[15,17,24,25] The Rasch model enabled us to convert the non-linear raw data into linear logit scores[25,26] which were then used to rank the visualisation skills in order of difficulty (Figures 2 and 3). We also used the logit scores to construct a person–item map (Figure 4) that indicated students' relative levels of VL.

Validation of the instrument

To improve the trustworthiness and credibility of our instrument, the lead researcher first developed the instrument independently by identifying relevant cognitive skills from Bloom's taxonomy. The list of identified skills was presented to the two co-researchers, who individually scrutinised the skills and probes for face and content validity. Concerns were negotiated among the researchers until consensus was reached. Thereafter an independent panel of experts was consulted to further validate the instrument for face and content validity. In line with a previous study[21], the panel was made up of nine experts, P1 to P9, including three postgraduate biochemistry students, two biochemistry lecturers (PhD-qualified) and four senior science educators.

The experts completed a questionnaire that required them to scrutinise each probe in terms of its legitimacy and appropriateness for the current study. In particular, they considered whether the probes were valid measures of each identified visualisation skill and the particular biochemistry propositional knowledge that was being assessed. The questionnaire consisted of both open-ended and closed-ended items. Responses to the closed items were scored on a Likert scale of 0, 1, 2 or 3 for 'strongly disagree', 'disagree', 'agree' and 'strongly agree', respectively. Content validity indices[21] were calculated to determine whether any of the designed probes required revision or exclusion. Based on the expert feedback, necessary adjustments were made to improve and optimise the test instrument.

Reliability of the instrument

The Rasch model was also employed to determine various measures of test instrument reliability. The first measured variable was 'dimensionality', which is an important measurement in using the Rasch model.[15,16,22] Unidimensionality was assumed if the fit statistics ranged between 0.5 and 1.7, as suggested by Velozo et al.[16] and Wright and Linacre[18]. The dimensionality of our items (visualisation skills) ranged from 0.56 to 1.6 for infit statistics, and from 0.58 to 1.66 for outfit statistics. These values suggested our data were unidimensional, which further justified the use of the Rasch model in our study.

Test–retest reliability was also calculated, to ensure internal consistency in scoring the items. For this assessment, the test was administered twice to one group, over 8 weeks. Internal consistency was measured using SPSS to determine if the mean scores were the same. The retest results indicated a statistically significant correlation between the initial and second tests, with a correlation coefficient of 0.45 ($p<0.0001$, 95% CI: 0.27 to 0.59). The results yielded a Cronbach alpha value of 0.798. These values showed that the results from the first test were consistent with the results of the second test, which indicates satisfactory internal consistency.

Administration of the instrument

Ethical clearance was obtained from a South African university (name withheld for ethical reasons; ethical clearance number HSS/0150/07). All participating students were over 18 years of age and gave their written consent to participate in the study. The test was administered to 106 third-year students, of whom 31 attended Campus A and 75 attended Campus B at the same university. The students were selected using a purposive (non-probability) sampling method[23]. All participating students were enrolled in a course on protein structure and function, which covered the propositional knowledge required to answer the probes.

We used predetermined and validated criteria to score student responses to the probes. The probes focused primarily on students' visualisation skills and knowledge of symbolic language, demonstrated by their ability to explain the concepts represented by symbols and graphical cues composing the ERs. The content of each ER was kept simple so that student responses were less influenced by conceptual knowledge and more by cognitive ability to interpret ERs.

Results

Reliability coefficients were calculated using the Rasch model to measure *test reliability* as well as *person reliability*. The reliability coefficients were therefore computed for both the Campus A and Campus B student groups. However, probably due to unequal student sample sizes, the Campus B data reflected a higher item reliability coefficient ($r = 0.96$) than that of the Campus A sample ($r = 0.93$). This was also reflected in the person reliabilities, which were 0.86 and 0.80 for the Campus B and Campus A data, respectively. Based on this analysis we concluded that the probes were reliable and the persons were reliable.

Ranking visualisation skills by level of difficulty

As indicated above, WINSTEPS Rasch software was used to generate the order of visualisation skill difficulty for both campuses as as shown in Figures 2 and 3. These figures show the order of difficulty of the visualisation skills (Table 1) as tested through the 12 probes. The *t*-test showed that difficulty levels for the general items (visualisation skills) did not differ significantly for students at the two campuses ($p>0.05$). However, we noted that the difficulty indices tended to change relative to standard deviations. As shown in Figures 2 and 3, the difficulty level of visualisation skills varied among students, such that overall each skill had a unique difficulty index. However, for both samples (i.e. both campuses), skill T02 ('arrange, order, organise, classify') was the 'easiest' and T18 ('perceive luminance, identify colours') was the 'most difficult'.

To obtain a more uniform order of item difficulty, data from both sample groups were combined to form one larger dataset. This approach was based on Linacre's[27] suggestion that in order to normalise or calibrate a scale of item difficulty, the sample size should range from 16 to 36 and from 27 to 61, such that the standard deviation may lie within $\pm$ 1 logit at a 95% and 99% level of confidence, respectively. In accordance with this approach, we used the combined linear logit scores to generate an item difficulty map (Figure 4).

The right hand column of Figure 4 indicates the order *and* level of difficulty of the 24 visualisation skills (as shown in Table 1). Skills that had similar scores (i.e. differed by less than 0.1 logit) do not have a separating line (|) to the left of the items. For instance, visualisation skills T07 ('Describe, discuss, explain') and T24 ('use') had scores that were similar. A similar pattern was observed with T08, T12, T13 and T23. The item mean (*M*) was set automatically by the Rasch model at zero, representing the average difficulty level, where items above 0 are more difficult and items below 0 are less difficult[15]. Letters *S* and *T* indicate 1 and 2 standard deviations from *M*, respectively.

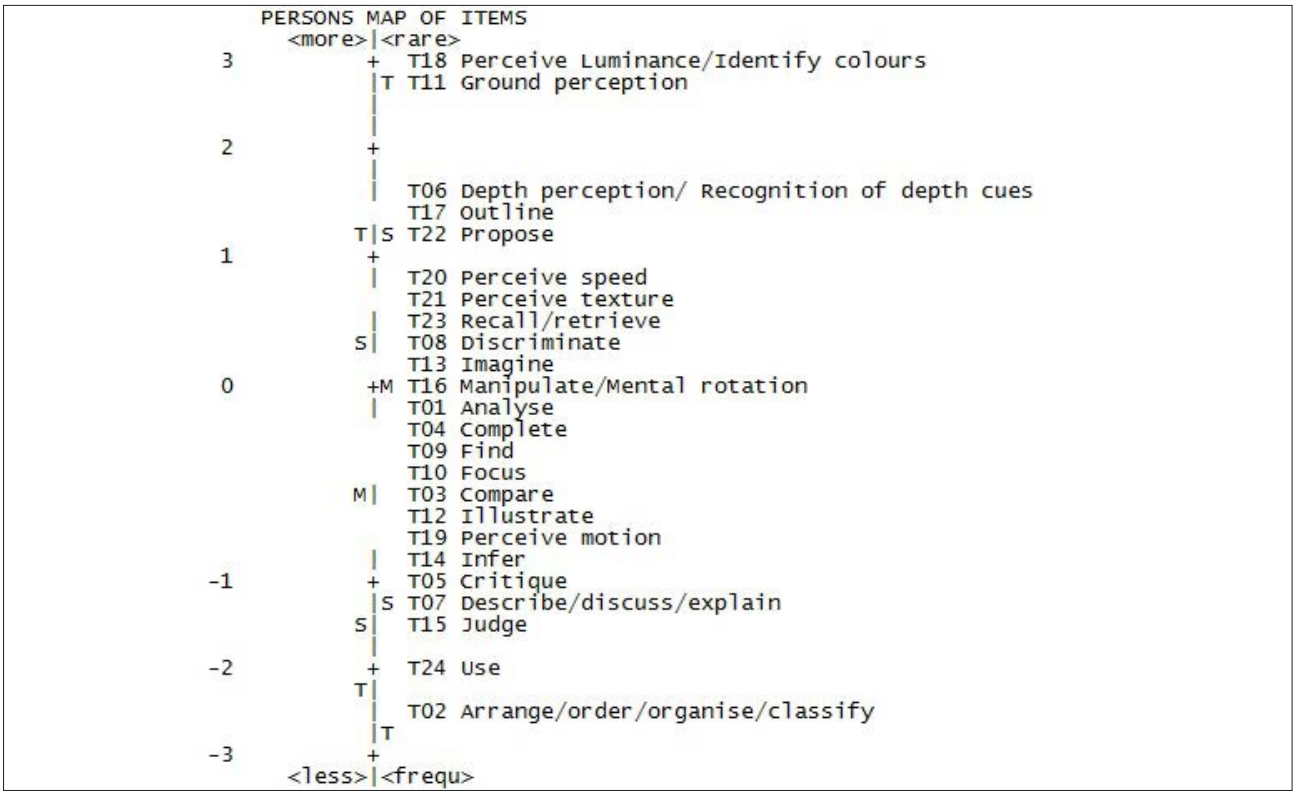


Figure 2: Item difficulty map for Campus A students (n=31).

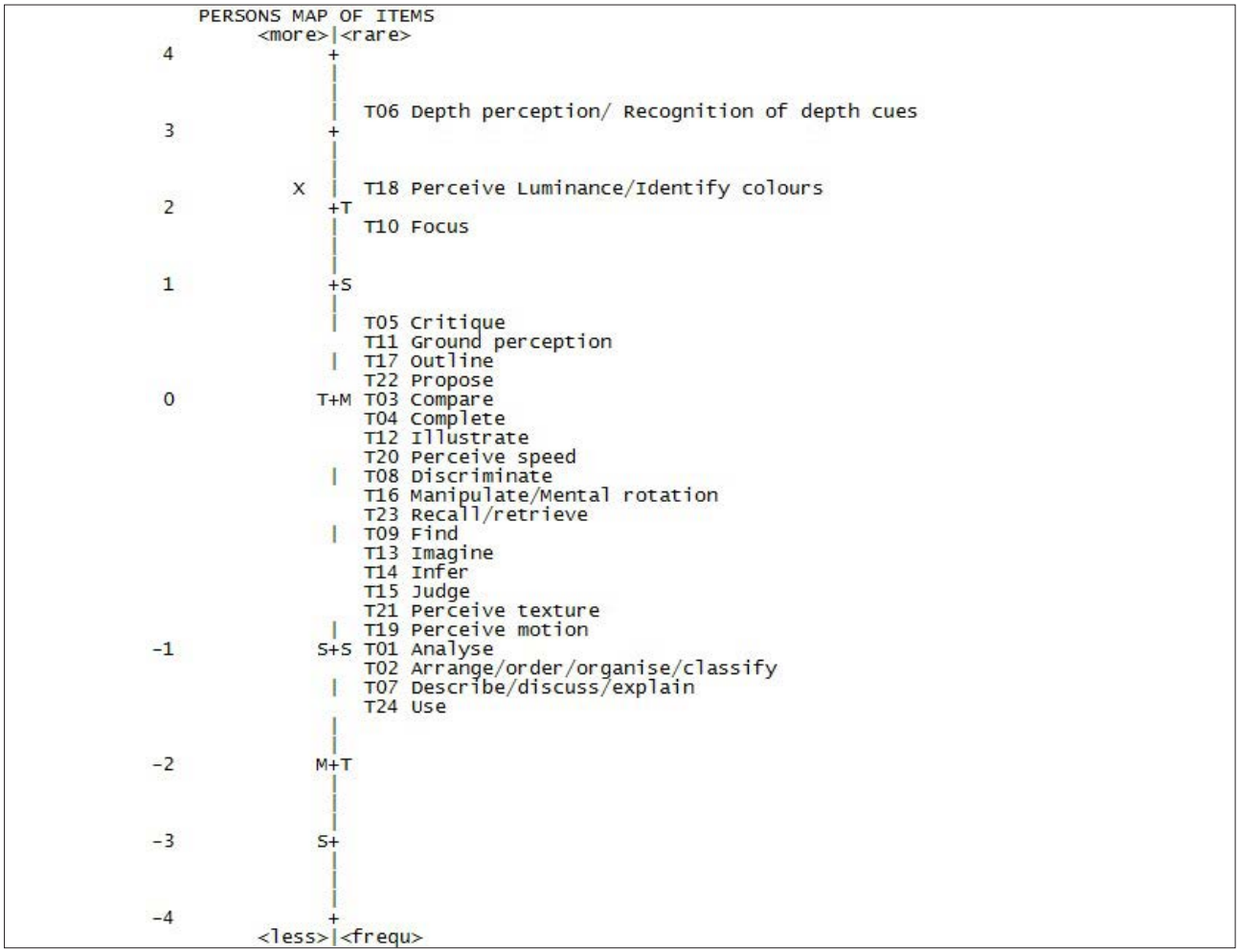


Figure 3: Item difficulty map for Campus B students (n=75).

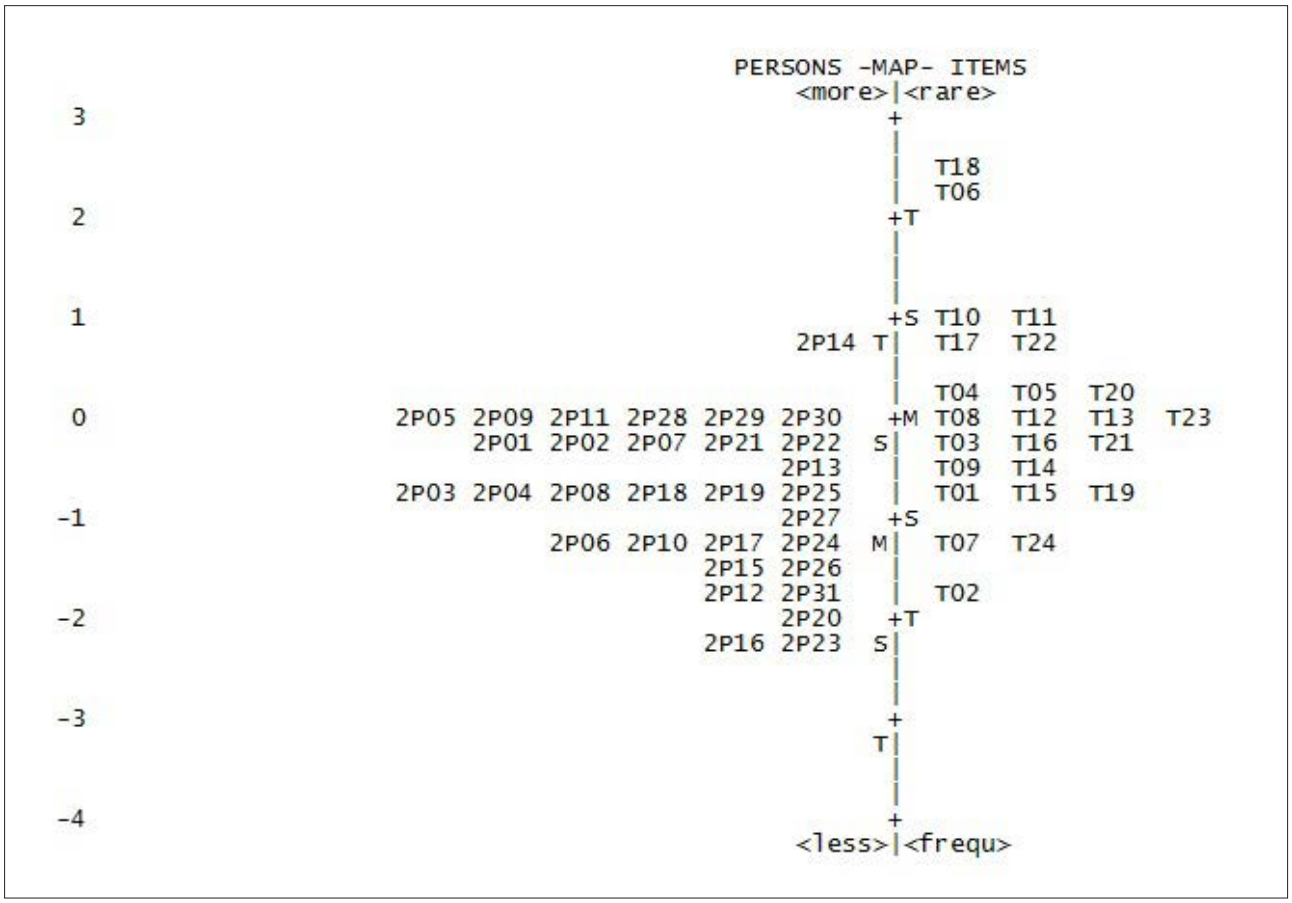


Figure 4: Person–item map showing Campus A students' level of visual literacy plotted against order of difficulty of visualisation skills. The left-hand side shows level of visual literacy for students from Campus A; the right-hand side shows order and level of difficulty for the 24 visualisation skills.

Determining relative levels of visual literacy

The retest results for the Campus A sample ($n=31$) were used to determine students' levels of VL. Using the Rasch analysis, we generated a person–item map for these 31 students (left side of Figure 4). The map shows person ability measures (i.e. Campus A students' VL levels) plotted on the same scale as item difficulty measures (order of difficulty for visualisation skills for the combined Campus A and B dataset). The person–item map (Figure 4) shows the students' level of VL in relation to the visualisation skills item map. The right-hand portion of Figure 4 shows the order and level of difficulty for the 24 visualisation skills (listed in Table 1). The left-hand portion of the person-item map displays a ranking of the participating students in accordance with their converted logit scores. The scale ranges from +3 to –3 logits, with the highest value corresponding to the most visually literate and most difficult skill, and the bottom of the scale indicating the least visually literate and least difficult skill.

The following inferences can be made from the person–item map (Figure 4). Firstly, students 2P05, 2P09, 2P11, 2P28, 2P29 and 2P30, who achieved a VL level of 0 logits, would have a 50% odds of correctly answering the probes for the corresponding visual skills (namely T08, T12 and T13)[15]. The same students would have only a 25% odds of correctly answering probes for visualisation skills T10 and T11, which have a difficulty level of 1 logit for those students. Secondly, the person–item map suggests that almost all participating students did not show an ability to perform skills T18, T06, T10, T11, T17, T17, T22, T04, T05 and T20 (see Table 1). Students were generally able to perform other skills. Thirdly, for visualisation skills as a whole, student 2P14 scored the highest, whereas students 2P16 and 2P23 scored the lowest in terms of person measure. These three students demonstrated the highest and lowest levels of VL in the Campus A class.

Notably, both the student and item scores followed an (overlapping) normal distribution, with an acceptable spread of values and match between the range of item and person measures, and there were few gaps in the data. This suggests that the test instrument was sensitive to differences in skill difficulty for the participating students, and was also sensitive to differences in student competence in these skills. Only two test items (T06 and T18) showed a difficulty level greater than 2 logits (i.e. greater than 2 standard deviations from *M*), and none of the person measures showed VL levels above 1 logit. Furthermore, *M* of person abilities (left portion of Figure 4) was about 1.25 logits lower than *M* for item measures (right portion), which suggests that the test instrument was generally too difficult for most students in the Campus A sample. This in turn suggests that either the standard of the visualisation test needs lowering, or that action needs to be taken to improve the visual competence of this group of students.

Regarding the need to improve the VL of biochemistry students, it is important to note that out of the 31 students in the Campus A sample, only one student (2P14) obtained a score above the item mean score. In other words, the remaining 30 students had a more than 50% chance of failing or scoring below average[15] on the VL test under study. Furthermore, three students (2P20, 2P16 and 2P23) scored below the lowest scoring visualisation skill (T02), and thus had more than 50% odds of failing all the visualisation skill probes in the test instrument. Thus, clearly the test instrument is very sensitive to differences in student levels of VL and could be used as a useful tool to inform both students and instructors about specific remedial needs to improve students' VL.

Discussion and conclusion

Research shows that a lack of VL is one of the major factors leading to poor content understanding among students.[7] Schonbörn and Anderson[20] also show that visualisation skills are necessary for students to comprehend content knowledge presented in ERs. Consequently, to better understand the causes of poor VL, we wanted to identify and assess visualisation skills required for students to effectively process ERs of biochemistry content. Our study identified and assessed 24 cognitive skills that can be regarded as visualisation skills for biochemistry. Therefore, in addition to teaching content knowledge, biochemistry education should foster the development of these specific visualisation skills so that students can better understand the complex and abstract biomolecular phenomena taught in biochemistry.

A key finding of our study is that VL varies among biochemistry students, as predicted in literature[3,8-10]. Determining students' levels of VL, and the levels of difficulty of specific skills, means that teachers and curriculum designers can make more informed decisions about prioritising the development of certain skills. This would address the problem identified by scholars, namely that students are not explicitly taught VL, probably because of a lack of suitable framework.[10] Furthermore, students' levels of VL can be measured so that developmental programmes can be designed to foster visualisation skills. In addition, instructors can make informed decisions regarding the choice of ERs to use, based on the presence or absence of visualisation skills among students.[3] For example, most students in our sample had difficulties with 'depth perception'. Therefore, ERs that require the application of this skill may prove challenging for most students. Instructors need to make sure that before utilising certain ERs, their students have the necessary skills required for optimal comprehension of content.

Schonbörn and Anderson[3] point out that the symbolic language used to represent and externalise biomolecular phenomena is difficult for students to master. We argue that the visualisation skills addressed in our study provide the vocabulary for this symbolic language. As the results of the Rasch model showed, students who had an average level of VL had a strong chance of developing and using visualisation skills below the average level. Therefore, the more skills students have, the higher they move on the VL scale, and the more adept they can become at communicating in the symbolic language of biochemistry.

Our study has contributed some unique findings to the literature. The main limitation was that the probes were tested on a sample of students who attended a single learning institution. There is a need to further calibrate the instrument through multiple rounds of testing with a broader sample. Further research is also required on approaches for improving the VL of students, and to remediate any visual difficulties and problems with visualisation skill competence where they might arise. In this regard, it is important to consider explicitly teaching visualisation (and other cognitive) skills as part of the formal biochemistry curriculum.[10]

Our study may stimulate further important research questions to improve the understanding of VL. For instance, educators and ER designers may ask questions such as: 'Besides conceptual knowledge, which visualisation skills are students expected to have in order to use a particular ER effectively?', 'What visualisation skill(s) is this ER promoting, and how?' and 'What types of ERs are appropriate for the educational level and visual competence of a certain group of students?' Asking such questions is crucial if we wish to see meaningful, effective learning and teaching take place with the use of ERs in biochemistry and the molecular life sciences.

Authors' contributions

L.M. was the main researcher, who led the work, collected and analysed the data, and prepared the manuscript. K.S. was the project co-supervisor and contributed to drafting the manuscript. T.A. was the supervisor of the project and contributed to drafting the manuscript.

References

1. Mnguni LE. The theoretical cognitive process of visualization for science education. SpringerPlus. 2014;3:184 http://dx.doi.org/10.1186/2193-1801-3-184
2. Burton L. Helping students become media literate. Workshop paper. Australian School Library Association (NSW) Inc. 5th State Conference; 2004.
3. Schonbörn KJ, Anderson TR. The importance of visual literacy in the education of biochemists. Biochem Mol Biol Educ. 2006;34(2):94–102. http://dx.doi.org/10.1002/bmb.2006.49403402094
4. Dori YJ, Barak M. Virtual and physics molecular modelling: Fostering model perception and spatial understanding. Educ technol soc. 2001;4(1):61–70.
5. Huk T, Steinke M. Learning cell biology with close-up views or connecting lines: Evidence for the structure mapping effect. Comput Human Behav. 2007;23(3):1089–1104. http://dx.doi.org/10.1016/j.chb.2006.10.004
6. Mayer RE, Hegarty M, Mayer S, Campbell J. When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. J Exp Psychol Appl. 2005;11(4):256–265. http://dx.doi.org/10.1037/1076-898x.11.4.256
7. Bottomley S, Chandler D, Morgan E, Helmerhorst E. jAMVLE, A new integrated molecular visualization learning environment. Biochem Mol Biol Educ. 2006;34(4):343–349. http://dx.doi.org/10.1002/bmb.2006.494034052666
8. Littlejohn A, Falconer I, Mcgill L. Characterising effective eLearning resources. Comput Educ. 2008;50(3):757–771. http://dx.doi.org/10.1016/j.compedu.2006.08.004
9. Grisham D, Power M, Riles P. Visual skills of poor readers in high school. Optometry. 2007;78(10):542–549. http://dx.doi.org/10.1016/j.optm.2007.02.017
10. Anderson TR, Schönborn KJ. Bridging the educational research–teaching practice gap: Conceptual understanding, part 1: The multifaceted nature of expert knowledge. Biochem Mol Biol Educ. 2008;36(4):309–315. http://dx.doi.org/10.1002/bmb.20209
11. Lowe R. Visual literacy and learning in science. [Document on the Internet.] ERIC Digests, Art. # ED463945. c 2000 [Cited 2012 February 20] Retrieved from http://www.ericdigests.org/2003-1/visual.htm.
12. Seufert T. Supporting coherence formation in learning from multiple representations. Learn Instr. 2003;13(2):227–237. http://dx.doi.org/10.1016/S0959-4752(02)00022-1
13. Sims E, O'Leary R, Cook J, Butland G. Visual literacy: What is it and do we need it to use learning technologies effectively? Paper presented at Australasian Society for Computers in Learning in Tertiary Education (ASCILITE). Auckland, New Zealand; 2002.
14. Anshel MH, Weatherby NL, Kang M, Watson T. Rasch calibration of a unidimensional perfectionism inventory for sport. Psychol Sport Exerc. 2009;10(1):210–216. http://dx.doi.org/10.1016/j.psychsport.2008.07.006
15. Bond TG, Fox CM. Applying the Rasch model. New Jersey: Lawrence Erlbaum; 2001.
16. Velozo CA, Kielhofner G, Gern A, Lin F-L, Azhar F, Lai J-S, et al. Worker role interview: Toward validation of a psychosocial work-related measure. J Occup Rehabil. 1999;9(3):153–168. http://dx.doi.org/10.1023/A:1021397600383
17. Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. BMC Med Res Methodol. 2008;8(1):33. ISSN 14712288 http://dx.doi.org/10.1186/1471-2288-8-33
18. Wright BD, Linacre JM, Gustafson JE, Martin-Löf P. Reasonable mean-square fit values. Rasch Measurement Transactions. 1994;8:370.
19. Anderson LW, Krathwohl DR, Airasian RW, Cruikshank KA, Mayer RE, Pintrich PR, et al., editors. A taxonomy for learning, teaching, and assessing. New York: Addison Wesley Longman; 2001.
20. Schonbörn KJ, Anderson TR. A model of factors determining students' ability to interpret external representations in biochemistry. Int J Sci Educ. 2009;31(2):193–232. http://dx.doi.org/10.1080/09500690701670535
21. Hyrkäs K, Appelqvist-Schmidlechner K, Oksa L. Validating an instrument for clinical supervision using an expert panel. Int J Nurs Stud. 2003;40(6):619–625. http://dx.doi.org/10.1016/S0020-7489(03)00036-1

22. Hsueh IP, Wang WC, Sheu CF, Hsieh CL. Rasch analysis of combining two indices to assess comprehensive ADL function in stroke patients. Stroke. 2004;35(3):721–726. http://dx.doi.org/10.1161/01.STR.0000117569.34232.76

23. Maree K. First steps in research. Pretoria: Van Schaik Publishers. 2007.

24. Boone W, Rogan J. Rigour in quantitative analysis: The promise of Rasch analysis techniques. Afr J Res Math Sci Technol Educ. 2005;9:25–38.

25. Pesudovs K, Garamendi E, Keeves JP, Elliott DB. The activities of daily vision scale for cataract surgery outcomes: Re-evaluating with Rasch analysis. Invest Ophth Vis Sci. 2003;44(7):2892–2899. http://dx.doi.org/10.1167/iovs.02-1075

26. O'Neill T. Definition of a logit. NCLEX Psychometric Technical Brief. 2005;2:1–3. Retrieved from https://www.ncsbn.org/02_18_05_brief.pdf

27. Linacre JM. Sample size and item calibration stability. Rasch Measurement Transactions. 1994;7(4):328.

Pesticide management practices among rural market gardening farmers near Harare, Zimbabwe

AUTHORS:
Moses Zimba[1]
Clemence Zimudzi[1]

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

CORRESPONDENCE TO:
Moses Zimba

EMAIL:
mzimba@science.uz.ac.zw

DATES:
Received: 11 Dec. 2015
Revised: 23 Feb. 2016
Accepted: 19 Apr. 2016

KEYWORDS:
organophosphates; pyrethroids; poisoning; counterfeit; agrochemicals

HOW TO CITE:
Zimba M, Zimudzi C. Pesticide management practices among rural market gardening farmers near Harare, Zimbabwe. S Afr J Sci. 2016;112(9/10), Art. #2015-0443, 5 pages. http://dx.doi.org/10.17159/sajs.2016/20150443

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

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

In 2014, we carried out a survey in Chinamhora and Chihota communal lands on the outskirts of Harare city, with the aim of understanding pesticide management practices among market gardening farmers. The farmers grew vegetables that mostly included tomatoes, cabbages, rape, cucumbers, onions and carrots, and they used mainly organophosphates and pyrethroids to control pests. A questionnaire was administered to 119 male heads of households across both study areas. The questionnaire contained 13 closed-ended questions in three sections: source and quality of pesticides, handling and use, and storage and disposal of pesticides used to protect crops. The study identified numerous gaps related to the handling of pesticides. Although the quality of labelling and packaging can largely identify the quality of pesticide, most of the farmers (77.3%) could not distinguish between genuine and counterfeit pesticides; approximately half (47.9%) of the farmers were not concerned about expiry dates; 27% did not observe post-spray periods; and 63% did not take precautions according to colour-coding of the pesticides. Also of concern were the large numbers of farmers who were not using protective coveralls (54.3%); a substantial number who were not using knapsacks for spraying (21.8%); poor storage of the pesticides, as shown by the variation in storage facilities; the use of empty pesticide containers for domestic purposes (20.2%); and lack of strict adherence to recommended dose levels, with some farmers (28.6%) merely estimating the dilution of pesticides. Training through outreach programmes is recommended.

Significance:

- Identifies gaps in the way pesticides are used and stored by rural market gardening farmers.
- Highlights the need for government agricultural extension workers to hold regular workshops for farmers.
- Indicates the need for government ministries to monitor counterfeit pesticides.

Introduction

Market gardening of vegetables in Zimbabwe involves intensive use of pesticides to maximise yields. Pesticides are inherently toxic materials, and have been an essential component of insect pest control since the early 1950s when organochlorine insecticides were first widely introduced.[1] Official records of the quantity of insecticides used in Zimbabwe annually are not available, but the amounts must be huge because the country has an agro-based economy. According to the World Health Organization (WHO[2]), the level of safe pesticide management in developing countries is low. When improperly used, pesticides can have effects such as poisoning through direct ingestion or accumulation in food, and may lead to insecticide resistance in target pests.[3]

The WHO[2] has defined pesticide management as the regulatory control, proper handling, import, supply, transport, storage, use and disposal of pesticide waste to minimise adverse environmental effects and human exposure. In Zimbabwe, it is a statutory requirement for pesticide containers to be marked with a triangle that is green, amber, red or purple based on the toxicity of the pesticide.[4] Green means formulations that have an acute oral dosage (LD50) of over 2001 mg/kg of bodyweight. These pesticides can be used without danger in the home, or where stated, as admixture to grain or other stored produce for human or animal consumption. They can be sold by any shop or store. The word 'Caution' is written in a green triangle, and 'Harmful if swallowed' appears at the base of the triangle.[4] An amber triangle means a formulation having an LD50 of between 501 and 2000, which can be used without danger in home gardens and for external use about the home. The word 'Danger' appears with a symbol of skull-and-crossbones within the amber triangle, and the word 'Poison' is written at the triangle base.[4]

Red formulations have an LD50 of between 100 and 500. Their use should generally be restricted to horticultural, agricultural, health or industrial pest control operations. They may be sold by a licensed dealer if part of the premises is set aside for the storage and sale of dangerous substances. The word 'Danger' with a skull-and-crossbones symbol appears within the red triangle, with the words 'Dangerous poison' beneath the base of the triangle.[4] Purple triangles mean a formulation with an LD50 of up to 1000. These pesticides may be sold only to persons whose business, profession or trade requires them. They may only be offered for sale by licensed dealers where part of the premises is set aside for the sale of dangerous substances. The dealer must keep a person register of all sales of this group of pesticide, each sale being countersigned by the purchaser or his nominee and the firm's license number noted. The word 'Danger' with a skull-and-crossbones symbol appears within the purple triangle, and 'Very dangerous poison' is written at the triangle base.[4]

In common with other countries, Zimbabwe has statutory instruments under the Environmental Management Act that regulate pesticide use, both in agriculture and for public health. Globally, the public health sector uses large quantities of pesticides.[5] However, the bulk of pesticides for public health are used indoors, for

example to control malaria vectors using indoor residual spraying, or to control bedbugs and cockroaches. By contrast, almost all agricultural pesticides are applied in the open environment. Monitoring and enforcement of regulations on the use of pesticides is a major challenge. According to WHO[2], poor capacity to enforce regulations leads to the excessive and unsafe use of pesticides, which can result in the contamination of food, drinking water and the environment, as well as affecting aquatic organisms. Globally, use of pesticides in agriculture has been reported to contribute to insecticide resistance in insect vectors of disease.[5] The availability of substandard, illegal and counterfeit pesticide products on the market is also of great concern.[1]

To promote pesticide management practices that are safe to human health and the environment, an International Code of Conduct on the distribution and use of pesticides was formulated.[2] A study on organophosphate poisoning in Zimbabwe from 1995 to 2000[6] showed that of the 92 children (under 10 years old) who ingested this type of pesticide, 62% of cases were the result of accidental ingestion. It is important to be able to read and understand the instructions on pesticide containers. Cases of misuse of pesticides resulting in death are common throughout the world. In the U.S., Morgan et al.[7] reported 135 cases and 8 deaths among people who ate watermelons that had been treated with albicarb, a systemic pesticide not registered for such use. Rowley et al.[8] reported an outbreak of eldrin poison in Attock district of Pakistan, in which out of 194 people who were poisoned, 19 died. Numerous cases of pesticide poisoning and deaths have also been reported in Zimbabwe.[9,10]

The use of appropriate protective clothing, including face masks, is strongly recommended. Chavez et al.[11] reported the occurrence of optic nerve atrophy among people who had been exposed to methyl bromide. Proper storage of pesticide is also an important management issue. Cases have been reported of people easily accessing pesticides for purposes of committing suicide, and those who survived suffering from chronic fibrotic changes.[12] In Zimbabwe, studies on occupational hazards of pesticide use and handling have shown that more than 50% of farm workers were exposed to organophosphates during spraying.[13]

The demand for market gardening produce to feed a rapidly growing urban population has resulted in intensive farming around urban centres in Zimbabwe. Pesticide management practices in these farming areas are little understood. We aimed to determine pesticide management practices among market gardening communities on the outskirts of Harare, with a view to promoting outreach training programmes. The specific objectives were to identify the pesticides used in market gardening by small-scale market gardening communities on the outskirts of Harare, who supply produce to Harare city; to examine pesticide storage, usage and disposal practices among the communities; and to recommend strategies and action plans to strengthen pesticide management practices among the communities.

Materials and methods

Study sites

The study was conducted in Chinamhora (17°06´S, 31°13´E) and Chihota (18°05´S, 31°06´E) communal areas, situated approximately 30 km and 27 km north and south-east of Harare respectively. In selecting the study sites, consideration was given to similarities in their physical and socio-economic characteristics. Communal farmers in both study areas practised intensive perennial commercial market gardening of vegetables, tomatoes, cucumbers and carrots, along the banks of slow-flowing perennial streams. The farmers also keep cattle and goats. Harare is the major market for their produce, where it is sold wholesale. The farmers we interviewed used agricultural pesticides intensively.

Participants were identified from records of market-gardening farmers kept by Harare Municipality at the Department of Community Services. The two study areas included 142 household, 78 in Chinamhora and 64 in Chihota. Of these households, 62 were sampled from Chinamhora and 57 from Chihota, resulting in a total of 119 sampled households. Most of the farmers had houses with brick walls under tiles or asbestos. All surveyed households had blair toilets. Because of their proximity to Harare, the two study sites have had both primary and secondary schools for decades, and this influenced the literacy rate of inhabitants of the two study sites.

Survey methodology

A pilot study was conducted in 2014 using a closed-ended questionnaire adopted by WHO in a world survey on pesticide registration and management[2], with minor modifications. The questionnaire had 13 questions and was administered to 119 male heads of households who were engaged in market gardening. All respondents were interviewed at their homesteads and they provided individual consent to be interviewed. After the interviews, we gave the farmers pamphlets with information on proper procurement, safe use, storage and disposal of pesticide containers.

Analysis of responses

Data were analysed according to the method used by WHO[2], after the responses were captured on a Microsoft Excel spreadsheet. For analysis of responses to a particular question, the denominator was the number of farmers who responded to that question. Frequencies were converted to percentages to show the frequency distribution of response categories. The chi-square (χ^2) test was used to determine whether the observed differences in percentages among response categories were statistically significant at $p<0.05$.

Results and discussion

We interviewed 119 farmers, 62 (52.1%) of whom lived in Chinamhora and 57 (47.9%) of whom lived in Chihota. Pesticides used by the farmers were of two types, according to their formulations, namely emulsifiable concentrates (EC) and wettable powders. The two most commonly used pesticides were organophosphates, which are marked by red triangles, and synthetic pyrethroids, which are marked by green triangles. Organophosphates include lambda-cyhalothrin (50% EC) and fenvalerate (20% EC), and synthetic pyrethroids include malathion (50% EC), diamethoate (40% EC) and amitraz (20% EC). As mentioned earlier, 'red' pesticide formulations should be used only for horticultural, agricultural, health or industrial pest control operations.[4] 'Green' pesticides can be used without danger in the home or where stated as admixture to grain or other stored produce for human or animal consumption.[4]

Most of the farmers in our study (83 or 69.7%) reported that they obtained their market-gardening pesticides from urban retail shops. Slightly more than a tenth of the sample (17 or 14.3%) said they obtained pesticides from rural retail shops; a few (8 or 6.7%) bought pesticides from rural cooperative facilities; and several (11 or 9.2%) purchased from other sources, such as other local farmers. The sources of pesticides varied, with a significant difference ($\chi^2=171.3$, d.f.$=3$, $p<0.00001$) between sources in terms of the frequency of household responses to this question. The fact that the majority of farmers said they obtained their pesticides from urban retail shops might indicate that this is a relatively cheap source.

Most of the farmers (81 or 68.1%) were concerned to a major extent by substandard and/or counterfeit pesticide products. More than a quarter (28 or 23.5%) indicated moderate concern, and a few (10 or 8.4%) had minor concerns (Figure 1). There was a significant difference ($\chi^2=103.0$, d.f.$=2$, $p<0.00001$) between response categories for degree of concern about counterfeit pesticides, in terms of household percentages.

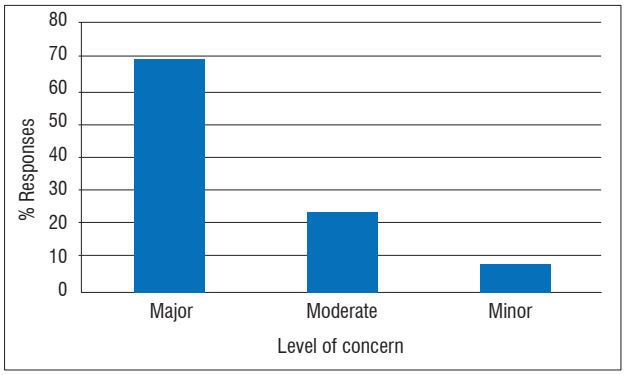


Figure 1: Response categories for concern about substandard or counterfeit pesticides.

Only 27 (22.7%) of the farmers we interviewed stated that they could distinguish between genuine and counterfeit pesticides to a large extent. A slightly bigger group (32 or 26.9%) said they could distinguish the difference to some extent, and a smaller group (20 or 16.8%) said they could not tell the difference at all. Roughly a third of participants (40 or 33.6%) said they could distinguish the difference to a small extent (Figure 2). There was a significant difference ($\chi^2=9.5$, d.f.$=3$, $p<0.05$) between response categories in percentage of household replies to the question about distinguishing between genuine and counterfeit pesticides.

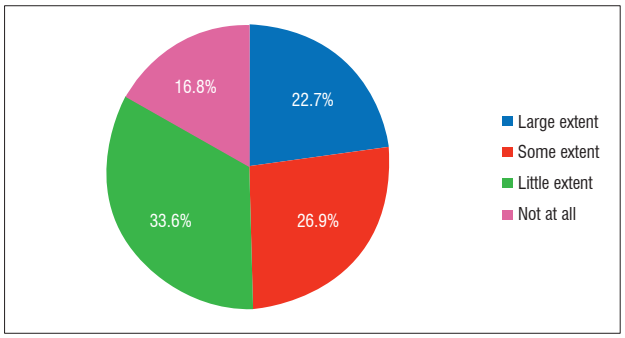


Figure 2: Responses about ability to distinguish genuine from counterfeit pesticides.

Because of the liberalisation of the economy, many unregistered pesticides have appeared on the market in Zimbabwe. A joint study by the Food and Agriculture Organization of the United Nations together with WHO showed that an estimated 30% of pesticides imported and marketed in developing nations (including in Africa) fall short of internationally accepted quality standards.[14] Our study showed that the ability of the farmers to distinguish genuine from counterfeit pesticides was poor. The use of substandard and counterfeit pesticide products has serious adverse effects on human health and the environment.[2] A recent survey involving 32 countries in the WHO Afro region showed that many countries were concerned by the trade and use of substandard pesticides.[2] However, despite these concerns, only 40% of African countries are reported to have national pesticide quality control facilities.[2]

All 119 respondents in our study reported that they were aware that pesticides have expiry dates. However, when asked whether they adhere to expiry dates, just over half (62 or 52.1%) of our respondents stated they always adhere to pesticide expiry dates; a smaller group (44 or 37%) said they adhere to pesticide expiry dates most of the time, and a tenth (13 or 10.9%) said they sometimes adhere to expiry dates (Figure 3). Hence, although a small majority of farmers stated they always adhered pesticide expiry dates, the large number of farmers who might not (47.9%) is cause for concern. There was a significant difference ($\chi^2=46.5$, d.f.$=2$, $p<0.00001$) between response categories in percentages of household replies to the question on adherence to expiry dates.

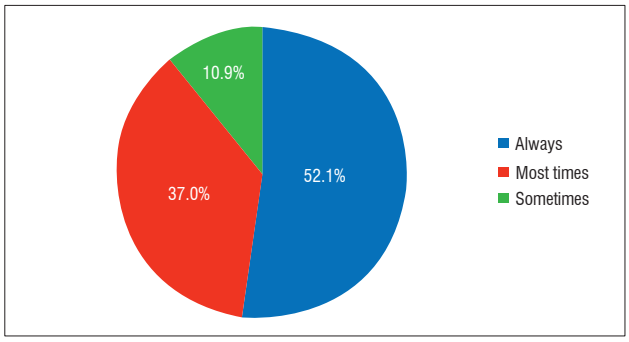


Figure 3: Responses about extent of adherence to pesticide expiry dates.

We also asked the participants about their adherence to a post-spray period before harvest. Overall, 73 (61.3%) of the farmers said they always observed a post-spray period, 30 (25.2%) observed a post-spray period most of the time, and 16 (13.4%) sometimes observed the period (Figure 4). Given that only 61% of the farmers said they always adhered to a post-spray period before harvest, evidently a large number of farmers in the study area were selling produce that still contained pesticides. There was a significant difference ($\chi^2=66.7$, d.f.$=2$, $p<0.00001$) between response categories in percentages of household replies to the question on observation of a post-spray period. Adherence to the recommended period before harvest is critically important to prevent poisoning through ingestion of pesticide residues.

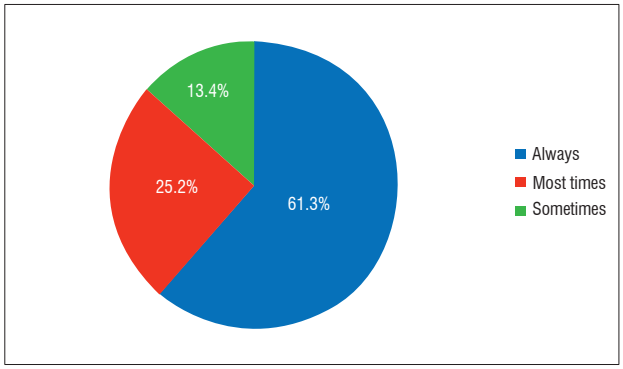


Figure 4: Responses about adherence to the recommended post-spray period.

With regard to colour coding, 44 (37.0%) of the farmers reported that they always take precautions according to the colour coding of pesticides. Just over a quarter (33 or 27.7%) said they do so most of the time, and the same number (33 or 27.7%) said they observed colour coding only sometimes; a minority (9 or 7.6%) said they do not observe colour coding at all (Figure 5). Less than 40% of respondents indicated that they always read instructions and take precautions according to the colour coding of a pesticide. There was a significant difference ($\chi^2=29.3$, d.f.$=3$, $p<0.00001$) between response categories in percentages of household replies to the question on precautions and colour coding.

A study among farm workers in Zimbabwe[15] showed that ignorance of colour codes was a major problem and a risk factor for pesticide exposure. Understanding colour coding on pesticide containers is therefore important to prevent pesticide poisoning. Labels carry colour codes to indicate the toxicity level of particular pesticides, and instructions on use and first aid information. In a similar study conducted among small-scale farmers in Nigeria[16], 88% of respondents said they did not refer to the material data sheets that accompanied pesticides. Cases of death from pesticide poisoning have been reported worldwide – for example a study by Morgan et al.[7] reported eight deaths among people who ate watermelons treated with albicarb. A study in Tanzania showed that most small-scale farmers in the study area applied pesticides that lacked specific instructions or labels, and 60% of respondents reported that they typically fell ill after pesticide application.[17]

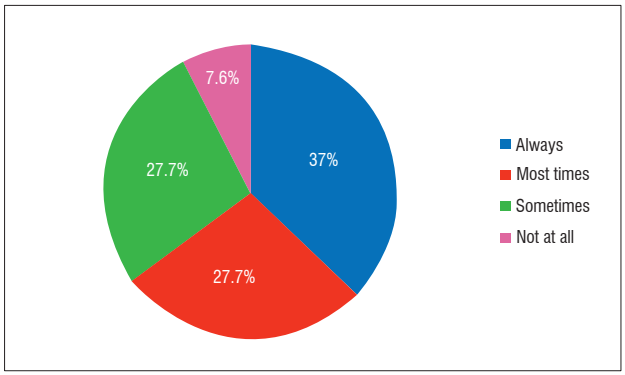


Figure 5: Responses about reading instructions and taking precautions based on colour coding of pesticides.

With regard to method of application, most farmers (93 or 78.2%) reported that they used knapsack sprayers; a tenth (13 or 10.9%) were using Hudson pressure pumps, and another tenth (13 or 10.9%) used traditional brooms. Knapsack sprayers and Hudson pressure pumps are the recommended pesticide application devices in Zimbabwe. Use of an appropriate device minimises exposure to and wastage of pesticides. Our findings showed that farmers in the study area do recognise the benefits of using the recommended equipment.

Fewer than half (52 or 43.7%) of the respondents said that they use coveralls such as overalls, work suits or dust coats when spraying. More than half (66 or 55.5%) stated that they use a respirator or piece of cloth to cover the nose, and more than half (70 or 58.8%) said they wear gumboots (Figure 6). There was a significant difference ($\chi^2=6.0$, d.f.$=2$, $p<0.05$) between response categories in terms of percentages of household replies to questions about protective clothing. The results showed that slightly more than half of the respondents used a face cover and gumboots, and fewer than half wore overalls, when using pesticides. Magauzi et al.[15] reported that low provision of protective clothing was a major risk factor for pesticide poisoning among farm workers in Zimbabwe. In many African countries, poor utilisation of protective clothing is a serious problem among small-scale farmers. Olowogbon[13] reported that 81% of workers did not use a complete outfit of recommended personal protective equipment while mixing and applying agrochemicals, and 79% did not calibrate the sprayers before use. These are major gaps in pesticide management, because protective clothing is meant to prevent entry of pesticides into the body.

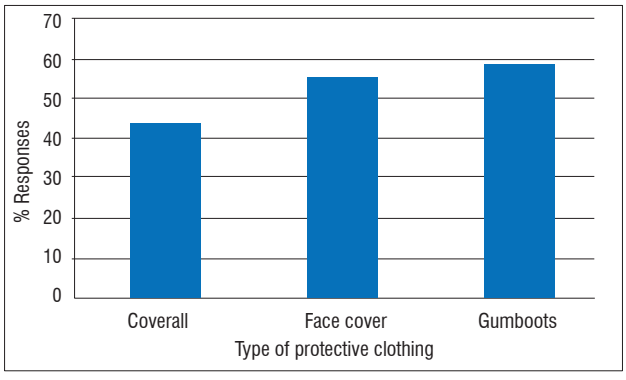


Figure 6: Responses regarding the use of protective clothing.

Only a quarter (30 or 25.2%) of the farmers reported that they had received formal training in the safe use of pesticides. A fifth of the sample group (24 or 20.2%) obtained information from the mass media; more than a third (45 or 37.8%) got their information from other farmers, and a sixth (20 or 16.8%) obtained information from the pesticide labels (Figure 7). There was a significant difference ($\chi^2=16.1$, d.f.$=3$, $p<0.05$) between response categories in percentages of household replies to the question about training in the safe use of pesticides. The low number of farmers who received training indicates that services by agricultural extension workers were not adequately reaching communal farmers. This finding confirms a WHO (2010) world survey, which highlighted gaps in training on safe use of pesticides both in government and the private sector.[2] The WHO study[2] also identified deficiencies in supervision, equipment maintenance, pesticide mixing and calibration in the countries surveyed.

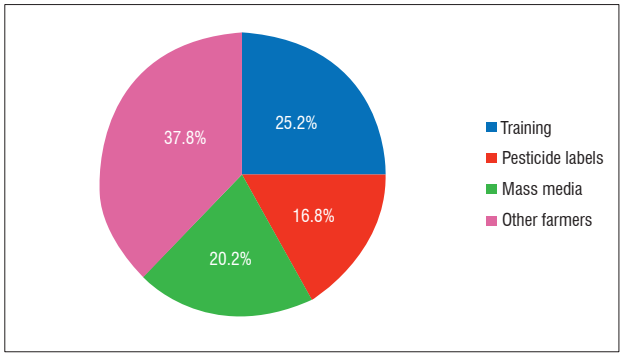


Figure 7: Responses about source of information on safe use of pesticides.

A third of the participants (39 or 32.8%) indicated that they used graduated containers to measure liquid pesticides when mixing. More than a third (46 or 38.7%) reported that they used containers they had calibrated, and just over a quarter (34 or 28.6%) relied on estimation.

With regard to storage, a quarter of the farmers (30 or 25.2%) stated that they stored pesticides in secure places in the field; just over a quarter (33 or 27.7%) said they kept pesticides in implement storerooms in the home; just over a fifth (26 or 21.8%) had chemical cabinets in the home; several participants (17 or 14.3%) stored pesticides in their granaries; and a few (13 or 10.9%) stated that they stored pesticides in the eaves of their houses (Figure 8). We statistically examined these variations in methods of pesticide storage. The results showed a significant difference ($\chi^2=15.3$, d.f.$=4$, $p<0.05$) between response categories in terms of household replies about the storage of pesticides.

The Food and Agriculture Organization[1] has highlighted the importance of rules on proper storage of pesticides in order to maintain product efficacy and to prevent contamination of the surroundings. The WHO[2] reported that storage of pesticides by small-scale farmers is still a major challenge in many developing countries. Studies on organophosphate poisoning in Zimbabwe from 1995 to 2000[6] showed that among cases of children (up to 10 years old) who ingested this group of pesticides, 62% cases involved accidental ingestion.[6]

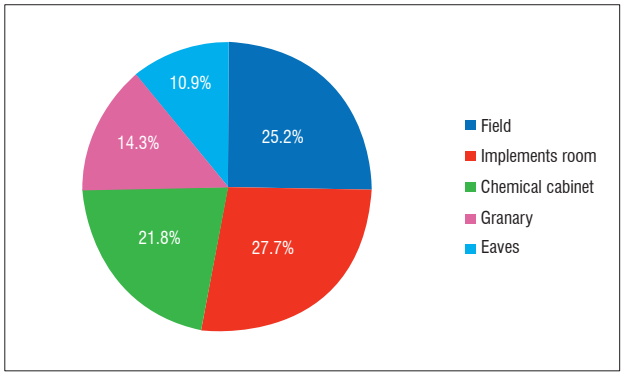


Figure 8: Responses regarding pesticide storage facility.

Almost half (55 or 46.2%) of the respondents stated that they repeated spraying until all mixed pesticide was finished. Slightly fewer participants (51 or 42.9%) said they kept excess mixed pesticides for the next round of spray. A few participants (5 or 4.2%) buried excess mixed pesticides in the ground, or gave it to a neighbour (8 or 6.7%) (Figure 9). There was a significant difference ($\chi^2=97.5$, d.f.$=3$, $p<0.00001$) between response categories in percentages of household replies about excess mixed pesticides.

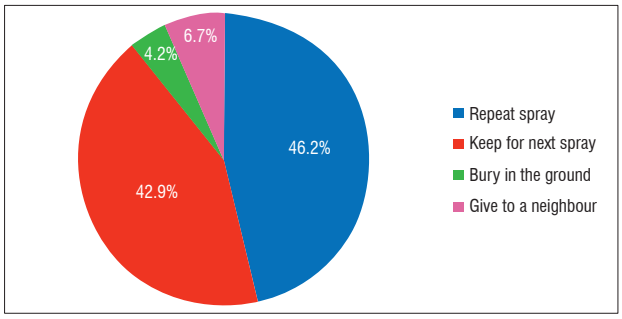


Figure 9: Responses regarding disposal of excess mixed pesticides.

Almost half of the farmers (54 or 45.4%) stated that they destroy all empty pesticide containers; a fifth (24 or 20.2%) clean all reusable empty containers and use them for domestic purposes; and slightly more than a third (41 or 34.5%) keep empty containers for pesticide mixing only (Figure 10). There was a significant difference ($\chi^2=17.1$, d.f.$=2$, $p<0.00001$) between response categories in percentages of household replies about disposal of empty pesticide containers.

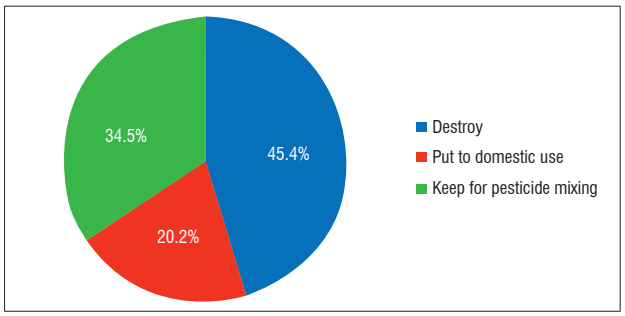


Figure 10: Responses regarding disposal of empty pesticide containers.

Safe disposal of pesticide waste, including used containers, is an important aspect of pesticide management to minimise risk to human health and the environment.[1] The vast majority of farmers in our study (89.1%) either said they continue spraying until all mixed pesticide is used or that they store it for the next round. Safe disposal of empty pesticide containers should be a policy of the ministry of health.[1] Our survey indicated that 20.2% of farmers convert to domestic use all usable empty containers after washing. Such use of pesticide containers for household purposes is hazardous, but it is common practice in many countries in the region.[18,19]

Recommendations

Our study has identified gaps in pesticide management practices among the rural market-gardening community on the outskirts of Harare. These gaps can be addressed only through education and training. There is a need for government agricultural extension workers to organise more outreach training programmes. We also recommend a survey to identify pesticide manufacturers and distributors, and the guidance they give to their rural farming clients in pesticide management. The training programme could take a multidisciplinary approach involving technical inputs and personnel from Ministries of Agriculture and the identified pesticide manufacturers and suppliers. The existing rural farmer training programme offered by agricultural extension officers could be reviewed to align it with current trends.

Authors' contributions

M.Z. initiated the study, conducted the survey, analysed the data and drafted the manuscript. C.Z. made conceptual contributions, conducted the survey and reviewed the draft manuscript.

Acknowledgements

We wish to thank Chinamhora and Chihota District Councils for permission to conduct the study.

References

1. Food and Agriculture Organization of the United Nations. International code of conduct on the distribution and use of pesticides [document on the Internet]. Rome: FAO; 2003 [cited 2015 March 10]. Available from: http://www.fao.org/docrep/005/y4544e/y4544e00.htm
2. World Health Organization. Public health pesticide registration and management practices by WHO member states [article on the Internet]. Geneva: WHO; 2011 [cited 2015 March 10]. Available from: http://www.whqlibdoc.who.int/publications/2011/9789241501217_eng.pdf
3. World Health Organization. Guidelines on public health pesticide management policy [document on the Internet]. Geneva: WHO; 2010 [cited 2015 March 15]. Available from: http://www.who.int/whopes/resources
4. Mudimu D, Chigume S, Chikanda M. Pestcide use and policy in Zimbabwe: A publication of the Pesticide Policy Project. Hannover: Uni Druck Publishers; Publication Series No 2; 1995. 47 pages.
5. World Health Organization. Global insecticide use for vector-borne disease control [document on the Internet]. Geneva: WHO; 2011 [cited 2015 March 21]. Available from: http://whqlibdoc.who.int/publications/2011/9789241502153_eng.pdf
6. Dong X, Simon MA. The epidemiology of organophosphate poisoning in urban Zimbabwe from 1995 to 2000. Int J Occup & Environ Health 2001;7(4):333–338. http://dx.doi.org/10.1179/oeh.2001.7.4.333
7. Green MA, Heumann MA, Wehr HM, Foster LR, Williams LP Jr, Polder JA, et al. An outbreak of watermelon-borne pesticide toxicity. Am J Pub Health. 1987;77:1431–1434. http://dx.doi.org/10.2105/AJPH.77.11.1431
8. Rowley DL, Rab MA, Hardjotanojo W, Liddle J, Burse VW, Saleem M, et al. Convulsions caused by endrin poisoning in Pakistan. Pediatrics. 1987;79:928–934.
9. Tagwireyi D, Ball DE, Nhachi CF. Poisoning in Zimbabwe: A survey of eight major referral hospitals. J Appl Toxicol. 2002;22(2):99–105. http://dx.doi.org/10.1002/jat.832
10. Tagwireyi D, Ball DE, Nhachi CF. Toxicoepidemiology in Zimbabwe: Pesticide poisoning admissions to major hospitals. Clin Toxicol (Phila). 2006;44(1):59–66. http://dx.doi.org/10.1080/15563650500394878
11. Chavez CT, Hepler RS, Straatsma BR. Methyl bromide optic atrophy. Am J Ophthalmol. 1985;99:715–719. http://dx.doi.org/10.1016/S0002-9394(14)76043-X
12. World Health Organization. Public health impact of pesticides used in agriculture [document on the Internet]. Geneva: WHO; 1990 [cited 2015 March 05]. Available from: http://www.who.int/iris/handle/10665/39772
13. Loewenson R, Nhachi CFB. Epidemiology of the health impact of pesticide use in Zimbabwe. In: Nhachi CFB, Kasilo OMJ, editors. Pesticides in Zimbabwe: Toxicity and health implications. Harare: University of Zimbabwe Publications; 1996. p. 25–35.
14. Food and Agriculture Organization / World Health Organization. Amount of poor-quality pesticides sold in developing countries alarmingly high [article on the Internet]. Rome: FAO/WHO; 2001 [cited 2015 March 12]. Available from: http://www.fao.org/WAICENT/OIS/PRESS_NE/PRESSENG/2001/pren0105.htm
15. Magauzi R, Mabaera B, Rusakaniko S, Chimusoro A, Ndlovu N, Tshimanga M, et al. Health effects of agrochemicals among farm workers in commercial farms of Kwekwe district, Zimbabwe. Pan Afr Med J. 2011;9:26. http://dx.doi.org/10.4314/pamj.v9i1.71201
16. Issa FO, Atala TK, Akpoko JG, Sanni SA. Adoption of recommended agrochemical practices among crop farmers in Kaduna and Ondo States, Nigeria. Journal of Agricultural Extension. 2015:19(1). 13 pages
17. Ngowi AVF, Mbise TJ, Ijani ASM, London L, Ajayi OC. Pesticides use by smallholder farmers in vegetable production in Northern Tanzania. Crop Prot. 2007;26(11):1617–1624. http://dx.doi.org/10.1016/j.cropro.2007.01.008
18. Food and Agriculture Organization / World Health Organization. Guidelines on management options for empty pesticide containers [article on the Internet]. Geneva: FAO/WHO; 2008 [cited 2015 March 25]. Available from: http://www.fao.org/fileadmin/templates/agphome/documents/.Pesticides/../Containers08.p
19. WHO. Safer access to pesticides: Community interventions [document on the Internet]. Geneva: WHO; 2006 [cited 2015 March 12.] Available from: http://www.who.int/mental_health/prevention/suicide/pesticides_safer_access.pdf

Do African microfinance institutions need efficiency for financial stability and social outreach?

AUTHORS:
Md A.K. Azad[1]
Susila Munisamy[1]
Abdul K.M. Masum[2]
Peter Wanke[3]

AFFILIATIONS:
[1]Department of Applied Statistics, University of Malaya, Kuala Lumpur, Malaysia
[2]Department of Administrative Studies and Politics, University of Malaya, Kuala Lumpur, Malaysia
[3]COPPEAD Graduate Business School, Federal University of Rio de Janeiro, Rua Paschoal Lemme, Rio de Janeiro, Brazil

CORRESPONDENCE TO:
Susila Munisamy

EMAIL:
susila@um.edu.my

DATES:
Received: 21 Dec. 2015
Revised: 10 Mar. 2016
Accepted: 28 Apr. 2016

KEYWORDS:
metafrontier; data envelopment analysis; productivity; world frontier technology

HOW TO CITE:
Azad MAK, Masum AKM, Munisamy S, Wanke P. Do African microfinance institutions need efficiency for financial stability and social outreach? S Afr J Sci. 2016;112(9/10), Art. #2015-0474, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20150474

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

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

Microfinance institutions (MFIs) have the dual objective of providing social welfare and financial stability. We evaluated the financial efficiency of MFIs in sub-Saharan African countries by comparing their regional performances during the period 2004–2013. We addressed prevailing MFI heterogeneity by using the concept of 'metafrontier'. The results showed that on an average, more than half the MFIs showed a drop in productivity. The measure of how much one country gets closer to or further away from world frontier technology is commonly known as the TGC score. In world frontier technology, East and South Asian countries have taken the lead (TGC score 1.0048) while sub-Saharan African countries lag behind (TGC score 1.0020). Most East and South Asian countries have a TGC score of 1, and most sub-Saharan African countries have a TGC score less than 1. This signifies that Asian countries lead world frontier technology and most African countries do not. The decomposition of efficiency scores showed that with regard to technical changes, African nations had progressed on average only 0.01%, and efficiency change scores had regressed by 0.59% annually.

Significance:

- First efficiency study on microfinance institutions and their heterogeneity in Africa.
- The results show robust discrimination among the efficiency scores.

Introduction

Microfinance institutions (MFIs) are commonly known as 'banks for the poor'. Mainstream financial institutions, such as commercial banks, do not allow poor households access to their services because of those households' poor economic status or creditworthiness. Thus, MFIs are designed to serve these people with a range of banking services, typically borrowing and saving. Most MFIs are very small in terms of capital to offer collateral-free credits. To support such financially risky business operations, MFIs usually depend on funds from donor organisations or individuals.

The two contradictory characteristics[1] of MFIs can be categorised as the 'institutional paradigm' (MFIs must meet their operational costs with financial sustainability) and the 'welfare paradigm' (social outreach).[2-4] In recent decades, MFIs have proven to be a successful tool for poverty alleviation and social outreach.[5,6] However, evidence of consistent poor loan quality among MFIs[7-9] has highlighted the pressing need to examine the financial sustainability of MFIs.

A number of studies have examined MFIs and their social performance in African nations.[10] To date, Van Rooyen and Stewart[11] have presented the most rigorous and systematic review containing quality data to assess the evidence of effectiveness of MFIs in sub-Saharan Africa (SSA). Their study showed that MFIs have modest but not uniform positive effects on social welfare. Hulme and Mosley[12] presented an extensive data comparison among the regions and showed that MFIs in Africa suffer from both micro and macro instability. Micro-level challenges include limited management ability and loan recovery rates, and macro-level challenges include inflation, interest rates and transparency. Hulme and Mosley[12] therefore could not rate the performance of MFIs as satisfactory. A similar finding was presented by Buckley[13], and contradictory findings were presented by Lafourcade and Isern[14]. All these studies attested to the performance of MFIs in poverty alleviation and social welfare. Our study examined MFI financial performance, especially in African countries, from 2004 to 2013.

In the literature, the topics of microfinance stability and institutional efficiency have come under the spotlight for two main reasons: (1) the conceptual difference between financial stability (i.e. end status) and efficiency (i.e. relative performance), and (2) increasing trends in operating expenses.[8] A decision-making unit (DMU) is seen as being efficient if the output cannot be increased without intensifying the input. In MFI efficiency literature, both regional[4,15,16] and country-wide[17-19] efforts are common. However, to date, few studies have examined the efficiency of MFIs globally. Annim[8] examined the efficiency of 164 MFIs worldwide. He employed parametric and non-parametric measures to evaluate MFI trade-offs between financial sustainability and social outreach. Hermes and Lensink[1] studied 435 MFIs worldwide using a stochastic frontier approach to measure both financial and social efficiency. Among these studies, a major limitation is their methodology, in that they did not properly account for heterogeneity among MFIs[20].

Worldwide, the heterogeneity of MFIs is evident for three reasons: geographical and socio-economic influence[21], the regulatory framework, and the institutional framework. For example, Guntz[21] reported that the average loan size of MFIs in East and South Asia (ESA) is USD149, whereas the average loan from MFIs in Eastern Europe and Central Asia is USD1579 – almost ten times as large. Based on legal and institutional frameworks, the sustainability scores of MFIs worldwide have been calculated by the Economist Intelligence Unit. Out of 55 selected countries, only 30 showed positive movement towards improvement in both regulatory and institutional frameworks for MFI operations. Nineteen countries were reported as having experienced negative change. Latin America and the Caribbean (LAC) had the highest overall regional score.

In a recent study, Louis et al.[22] studied trade-offs of MFIs with special attention to MFI heterogeneity. The researchers used self-organising map methodology to study 650 MFIs, and the results showed a positive relationship between MFI heterogeneity and social efficiency. This suggests that studying the efficiency and heterogeneity of African MFIs would also be worthwhile.

The concept of 'metafrontier' originated with Hayami[23] to deal with heterogeneity in efficiency calculations among DMUs. Hayami realised that studying efficiency on a comparative basis would be difficult because individual groups might not enjoy identical sets of production factors such as land, labour and capital. More information on meta-production functions is provided by Binswanger and Ruttan[24]. Their study showed that meta-production function enveloped all sub-functions, and was the most efficient function (assuming all groups had access to meta-production technology).[25] By modifying a single data-generation process, Battese and Rao[26] solved the drawback they had identified for using a stochastic frontier approach. In addition, the Malmquist Index (MI) was used to chart the yearly efficiency changes among selected MFIs.

Non-parametric data envelopment analysis (DEA) has been used as a tool to measure efficiency.[4,15,27] The main benefit of using DEA rather than parametric tests (e.g. stochastic frontier approach) is that DEA does not require any assumptions about DMU business processes. Moreover, DEA can examine efficiency using multiple inputs and multiple outputs. To the best of our knowledge, our use of metafrontier Malmquist DEA analysis is a first in examining MFI efficiency worldwide.

Our objective was to examine the efficiency of MFIs in Africa, with consideration given to heterogeneity in production technology. We classified all MFIs into five groups (Table 1) according to geographical and socioeconomic data, as proposed by earlier studies[8,21]. Metafrontier technology is used to handle such heterogeneity. We used the MI to capture efficiency changes over the study period. Then we decomposed MI values into three specific efficiency scores, namely technical efficiency, pure efficiency and scale efficiency, to examine the sources of efficiency results on the MI values. For the calculations, we used DEA considering the production approach of MFIs at variable returns to scale.

Microfinance worldwide

Poverty is a global issue. According to the World Bank[28] around 2.2 billion people currently live on less than USD2 a day, compared with 2.59 billion people in 1981. The report showed that poverty has declined slightly over the years, but not evenly across all regions. With regard to reductions in extreme poverty from 1981 to 2011, East Asia performed best (dropping from 78% impoverishment to a mere 8%), followed by South Asia (61% to 25%). Sub-Saharan Africa showed a less dramatic improvement, dropping from 53% impoverishment to 47%. In 2011, the largest extremely poor population was located in SSA (415 million people), followed by South Asia (399 million), and then East Asia and the Pacific (161 million).

Detailed differences in MFI operations across the regions are shown in Figures 1 and 2. Figure 1 shows the change in total number of MFI operations across the five regions. After making steady progress between 2000 and 2010, all regions faced a significant drop in the number of MFIs between 2010 and 2013. The largest drop was recorded for LAC and the smallest drop for Middle East and North Africa (MENA). Interestingly, for MENA the number of MFIs did not grow significantly over the years.

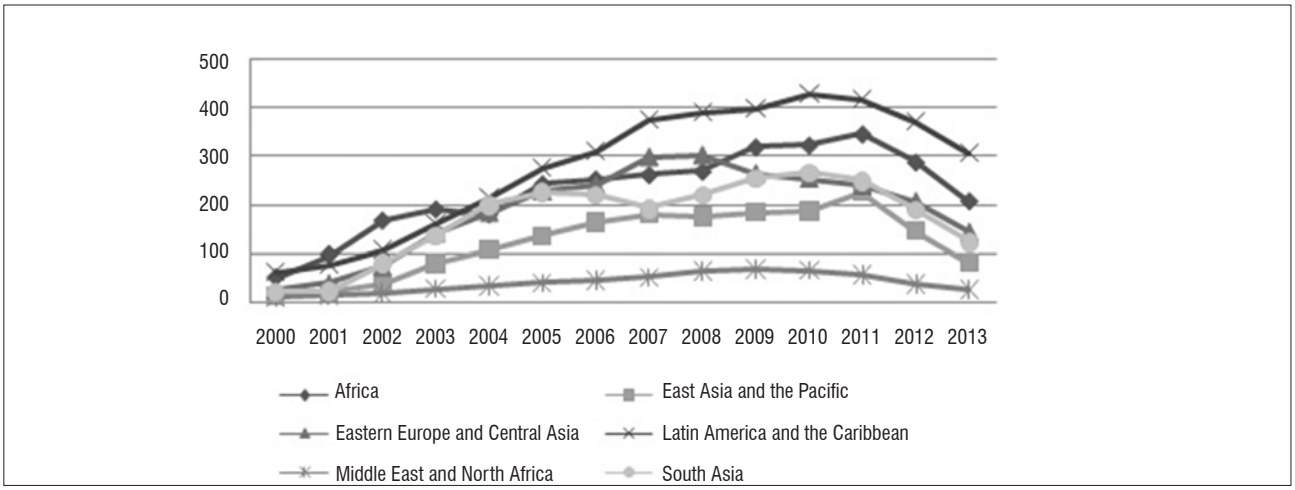


Figure 1: Number of microfinance institutions (2000–2013).

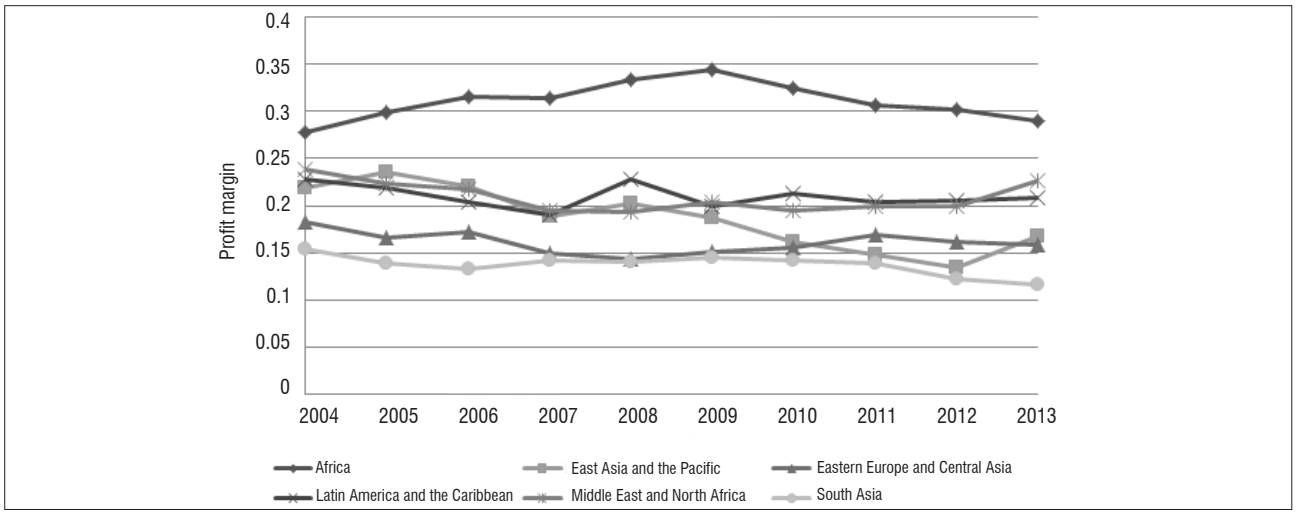


Figure 2: Profit margin trends among microfinance institutions (2004–2013).

Figure 2 presents a summary of financial stability among MFIs. The highest profit margin was recorded for Africa, whereas the lowest margin was in South Asia. After 2009, just after the global financial crisis, profit margin dropped in all regions but not substantially – except in Africa. Table 1 shows details about the heterogeneity among MFIs in five selected regions (country details, active borrowers, gross loan portfolio, Economist Intelligence Unit scores, and changes in scores from 2011 to 2013). From Table 1 it is evident that countries show unequal improvements in poverty alleviation. The greatest change was observed for LAC, and the smallest change was noted for ESA. Thus, MFI efficiency studies among these regions must consider their heterogeneity.

Previous studies on microfinance efficiency

Like other social organisations, MFIs operate under the 'double bottom line', that is, social outreach and financial stability.[3,5,8] Over the past few decades, MFIs have proven their success in social outreach.[1,6,28] However, until recently the financial sustainability of MFIs has not been thoroughly analysed.[21,29] According to Yunus[30], MFI financial sustainability can be seen as corresponding to social outreach, because establishing an institutional paradigm can ensure long-term operation and service to society. An efficient MFI can serve the welfare purpose better than a bankrupt MFI.

Table 1: Microfinance institutions worldwide

	MFIs	Active borrowers	Gross loan portfolio (USD)	EIU score[1]	Change[2]		MFIs	Active borrowers	Gross loan portfolio (USD)	EIU score	Change
Sub-Saharan Africa						**Latin America and the Caribbean**					
Cameroon	10	177,926	372,721,222	31.7	I	Argentina	9	24,536	40,070,698	28.8	NC
Congo	9	42,714	38,236,114	28.4	I	Bolivia	19	1,214,402	4,503,135,663	69.8	D
Ghana	8	158,530	106,348,688	53.3	I	Brazil	9	1,898,000	1,683,630,073	49.1	D
Kenya	17	1,084,136	2,590,799,699	61.1	D	Chile	3	272,275	1,896,443,997	49.1	D
Madagascar	7	120,802	108,542,288	35.9	NC	Colombia	18	2,588,526	7,610,872,312	58.5	I
Mozambique	3	34,775	37,182,502	44.0	NC	Costa Rica	14	17,689	74,010,673	42.1	I
Nigeria	11	1,184,776	351,272,355	48.2	I	Dom. Republic	12	429,801	716,656,089	53.6	I
Rwanda	20	125,193	433,603,375	48.4	D	Ecuador	47	1,441,065	3,815,893,767	48.3	D
Senegal	31	96,807	403,175,565	34.4	I	El Salvador	11	139,787	368,458,455	53.8	D
Tanzania	10	173,832	1,274,485,713	47.9	I	Guatemala	15	356,825	189,154,072	41.4	NC
Uganda	8	290,378	384,669,726	53.8	I	Haiti	4	123,199	59,013,056	25.8	D
Middle East and North Africa						Honduras	23	181,109	387,892,273	47.2	I
						Jamaica	2	20,846	22,057,946	31.8	I
						Mexico	51	6,618,779	2,710,689,954	51.1	D
Egypt	6	350,049	147,312,930	27.3	D	Nicaragua	22	316,024	313,034,496	52.9	I
Lebanon	3	47,405	60,353,727	33.3	D	Panama	5	41,056	164,200,973	53.5	D
Morocco	6	701,221	566,919,522	38.3	I	Paraguay	5	808,566	1,368,091,062	53.5	I
Yemen	2	44,409	8,860,768	31.0	I	Peru	44	4,858,582	10,038,700,349	82.5	I
East and South Asia						Trinidad & Tobago	1	3,585	559,138	26.5	I
						Eastern Europe and Central Asia					
Cambodia	16	1,826,845	2,637,058,170	60.3	I						
China	3	187,793	277,993,149	39.1	I	Armenia	12	239,801	872,110,229	47.4	NC
Indonesia	4	549,818	54,189,977	46.5	I	Azerbaijan	21	614,777	2,055,980,262	52.4	I
Philippines	20	2,477,724	535,703,157	67.9	I	Bosnia and Herzegovina	9	181,778	326,200,902	45.2	D
Vietnam	24	7,594,007	5,940,336,835	25.6	I	Georgia	10	228,685	831,679,818	43.4	I
Bangladesh	16	18,496,675	3,334,001,470	32.8	NC	Kazakhstan	8	165,339	461,200,347	36.0	D
India	88	32,545,085	5,471,886,863	62.0	I	Kyrgyzstan	9	291,906	344,466,610	35.1	D
Pakistan	15	1,532,769	346,586,244	69.7	I	Tajikistan	14	254,541	426,182,393	36.0	D
Mongolia	8	403,919	2,178,276,520	48.9	I	Turkey	1	65,637	19,988,400	26.5	D

[1]The EIU score is based on (1) regulatory framework and practices (regulation and supervision of microcredit portfolios, formation of regulated or supervised microcredit institutions, formation or operation of non-regulated microcredit institutions, regulatory and supervisory capacity for microfinance, and regulatory framework for deposit-taking); (2) supporting institutional framework (accounting transparency and client protection, achieved by transparency in pricing, dispute resolution, credit bureaus, and policy and practice for financial transactions through agents); and (3) adjustment factors, especially stability (political shock to microfinance and political stability). For details see *The 2013 Microscope Index and Report* at https://www.eiu.com/microscope2013 (cited 4/5/2015).

[2]I, increase or positive change (2011–2013); D, decrease or negative change (2011–2013); NC, no change (2011–2013).

Among the recent studies, Im and Sun[29] examined 1129 MFIs worldwide (in 98 countries) to test the institutional logic of MFIs. According to their study results, there is an inverted U-shaped relationship between financial stability and social outreach. Such a result signifies that MFIs must choose a point in profitability that maximises the MFI's potential for social outreach. After the threshold point, an increase in profitability will result in poor social welfare. Moreover, recent evidence of MFIs' declining financial performance[9] highlights the need to examine the efficiency of MFIs.

Traditionally, MFI efficiency is calculated using various accounting ratios. For example, Farrington[31] examined African MFI efficiency using cost-per-borrower and cost-per-sever ratios, and found that African MFIs are highly expensive. But these ratios explain only partial attributes of efficiency measurement, and as a result may misguide benchmarking and decision-making processes.[32] Among frontier efficiency studies of MFIs, the two main measurement techniques are non-parametric tests and parametric tests.[33,34] DEA is a non-parametric efficiency measurement technique developed by Charnes and Cooper[35]. DEA generalises the single input and single output measure of Farrell[32] into multiple input–output measures, to measure relative efficiency among the DMUs.[36] A DMU (in this context an MFI) is considered to be efficient if no DMU can produce equivalent outputs without needing increased inputs. One of the main benefits of using DEA rather than parametric efficiency techniques (the most popular one being stochastic frontier approach) is that DEA does not require detailed theoretical knowledge about processes.[37] In addition, the metafrontier DEA allows separation of DMUs based on their specific characteristics and thus the heterogeneity issue of DEA is resolved.[26]

In MFI efficiency studies using DEA, country-specific research effort is common.[17,38,39] Among many such studies, Tahir and Tahrim[17] examined efficiency of Cambodian MFIs using DEA and Dynamic MI. Their study showed that Cambodian MFIs were 92% efficient. With the help of MI, the decomposition of this result showed that efficiency was not achieved because of pure efficiency but rather scale efficiency. Therefore, their study confirmed that Cambodian MFIs were efficient in scale of operation and relatively inefficient at managing assets and costs. In a study of 46 Vietnamese MFIs, Nghiem and Coelli[39] deployed second-stage DEA to calculate technical efficiency and scale efficiency. Vietnamese MFIs were shown to be only 80% technically efficient. In the second stage of DEA, they used Tobit regression to examine sources of inefficiency. The findings showed that location of MFIs was the most significant variable in determining MFI efficiency. Wijesiri and Viganò[40], in a similar study in Sri Lanka, showed that MFIs were not efficient in either financial or social aspects. Using second-stage DEA with double bootstrap regression, Wijesiri and Viganò[40] showed that capital for assets, and age, were significant determinants of MFIs' financial efficiency.

To date, limited regional studies have also examined MFI efficiency. Bassem[16] evaluated the efficiency of 33 MFIs in MENA countries using DEA and MI. His study showed that MFIs are less scale efficient and are instead high in pure efficiency in the MENA region. This finding signifies that MFIs in MENA seem to be expert in management practices rather than optimum scale. Similarly, Servin and Lensink[41] examined 315 MFIs from 18 countries in LAC. Studies on cross-regional MFI efficiency are very few. For instance, Annim[8] examined only 164 MFIs worldwide to evaluate trade-offs between MFIs' social and financial performance. The results showed that worldwide, on average, MFIs failed at outreach service in terms of financial stability. Studying the cost efficiency of 39 MFIs in Africa, Asia and Latin America, Haq and Skully[4] showed that non-governmental MFIs are cost-efficient compared with bank microfinance. These studies, however, did not recognise heterogeneity among MFIs worldwide; this detracts from the robustness of the results[20]. Therefore a study on MFI efficiency globally, with proper consideration being given to their heterogeneity, is a timely need.

Recent research findings have revealed a significant importance of social performance of MFIs, over and above their financial efficiency. However, for a number of reasons the financial performance of MFIs can lead them to bankruptcy. Sainz-Fernandez et al.[7] examined crises among 832 MFIs in 74 countries from 2003 to 2011. The researchers analysed panel data and identified both internal and external reasons that were responsible for the MFI failures.

A proliferation of theories has created fuzziness in defining specifically what MFIs stand for.[41] One study that examined major trends in theories about MFIs showed that sustainable development was the most important aspect.[42] Another recent trend is 'green' MFIs. Allet[43] examined 160 MFIs using qualitative semi-structured interviews with 23 top managers, to understand why MFIs are deciding to go green. The results showed that stakeholders are the key drivers for this trend among MFIs.

Metafrontier Malmquist Index with DEA

Oh and Lee[44] introduced three technologies in metafrontier analysis: (1) contemporaneous distance function, (2) intertemporal distance function, and (3) global distance function. With the help of these benchmark technologies, we measured component distance function as required in metafrontier technology.

We assume that there are (J) different technologies within the selected DMUs. Contemporaneous benchmark technology produces a reference set (P) at an individual time period (t). For each group of technology (R_j), the production set is designed as $P^t_{R_j} = (x,y) \mid x \text{ produces } y$ and $\lambda P^t = P^t,\ t=1, \dots, T$, and $\lambda > 0$. This technology is based on the valuable work by Pastor and Lovell[45] and Tulkens and Vanden Eeckaut[46].

Tulkens and Vanden Eeckaut[46] also guided the second technology, namely intertemporal benchmarking. This is a simple combination of all $P^{InterT}_{R_j} = conv(P^1_{R_j} \cup P^2_{R_j} \cup, \dots, \cup P^2_{R_j})$ the proposed contemporaneous production sets and for all time periods for a defined technology group (R_j). So for all (J) different technologies within the selected DMUs, (J) different intertemporal benchmarks will be produced. Finally, global technology for all time periods is represented by $P^{Global}_{R_j} = conv(P^{InterT}_{R_1} \cup P^{InterT}_{R_2} \cup, \dots, \cup P^{InterT}_{R_J})$. Starting with the basic MI model of Caves and Christensen[47], a contemporaneous MI index would be as follows:

$$MI^3(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{D^3(x^{t+1}, y^{t+1})}{D^3(x^t, y^t)} \qquad \text{Equation 1}$$

where the production set is $P^s_{Rj}, s = t, t+1$ for R_j and the distance function is $D^s(x,y) = \inf\{\phi > 0 \mid \frac{x,y}{\phi} \in P^s_{R_j}$. Fare and Grosskopf[48] proposed the MI index as the geometric mean of MI of two periods because $MI^t(x^t, y^t, x^{t+1}, y^{t+1}) \neq MI^{t+1}(x^t, y^t, x^{t+1}, y^{t+1})$. With this connection, for an intertemporal benchmark technology the distance function is as follows:

$$MI^I(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{D^I(x^{t+1}, y^{t+1})}{D^I(x^t, y^t)} \qquad \text{Equation 2}$$

Here the production set is $P^I_{Rj}, I = t$ for a group of R^I_j and the distance function $D^I(x,y) = \inf\{\phi > 0 \mid \frac{x,y}{\phi} \in P^I_{R_j}$. Based on the valued work of Pastor and Lovell[45], any intertemporal distance function can be decomposed as follows:

$$MI^I(x^t, y^t, x^{t+1}, y^{t+1})$$

$$= \frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)} \times \left\{ \frac{D^I(x^{t+1}, y^{t+1})}{D^{t+1}(x^{t+1}, y^{t+1})} \times \frac{D^t(x^t, y^t)}{D^I(x^t, y^t)} \right\}$$

$$= \frac{TEff^{t+1}}{TEff^t} \times \frac{BPGp^{t+1}}{BPGp^t}$$

$$= EffC \times BpGCh \qquad \text{Equation 3}$$

where $TEff^s$ and $BPGp^s$ ($s = t, t+1$) represent the technical efficiency level and gap in technology for best practice, respectively. The term *EffC* denotes measure in change of efficiency as proposed by Fare and Grosskopf[48]. The term *BpGCh* denotes the changes in best practice technology gap between the contemporaneous and intertemporal production possibility frontier. The term $BpGCh > 1$ means that the contemporaneous frontier of $t+1$ period is closer than the intertemporal benchmark technology for the time t, and vice versa for $BpGCh < 1$. Pastor and Lovell[45] proposed this change in *BpGCh* is merely a change in the technology within the defined group. This is also the equivalent of technical progress or regress as presented by Caves and Christensen[47].

The metafrontier approach used in our study is defined in the production set of P^{Global} as follows:

$$MI^{Global}(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{D^{Global}(x^{t+1}, y^{t+1})}{D^{Global}(x^t, y^t)} \quad \text{Equation 4}$$

Here the production set is $P_{R_j}^{Global}$, $I=t$ for all groups of R_j^s, and the distance function $D^{Global}(x,y) = \inf\{\phi>0 \mid \frac{x,y}{\phi} \in P^{Global}$ is known as the global technology set. Using the same technology in Equation 1 for MI, the decomposition of a global set can be shown as follows:

$$MI^{Global}(x^t, y^t, x^{t+1}, y^{t+1})$$

$$= \frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)} \times \left\{ \frac{D^t(x^t, y^t)}{D^{t+1}(x^{t+1}, y^{t+1})} \times \frac{D^{Global}(x^{t+1}, y^{t+1})}{D^{Global}(x^t, y^t)} \right\}$$

$$= \frac{TEff^{t+1}}{TEff^t} \times \frac{BPGp^{t+1}}{BPGp^t} \times \frac{TGpR^{t+1}}{TGpR^t}$$

$$= EffC \times BpGCh \times TGpCh \quad \text{Equation 5}$$

Here, $TEff^s$, $BPGp^s$ and $TGpR$ ($s = t, t+1$) represent the technical efficiency level, technology gap for best practice, and the level of technological gap ratio, respectively. The technological gap ratio was introduced and empirically used by Battese and Rao[26]. This ratio identifies the gap between different technology groups used in sampling with the global technology set.

The distance function for $k' \in R_j$ for the period of $s = t, t+1$, is

$$(D^s(x^{k',s}, y^{k',s}))^{-1} = \max \phi_c^{k',s}$$

subject to:

$$\sum_{k \in Rj} z^{k,s} y_m^{k,s} \geq \phi_c^{k',s} y_m^{k',s}, \quad m=1 \ldots \ldots \ldots M$$

$$\sum_{k \in Rj} z^k x_n^{k,s} \leq x_n^{k',s}, \quad n=1 \ldots \ldots \ldots N$$

$$z^{k,s} \geq 0 \quad \text{Equation 6}$$

where z^k is the intensity variable of a DMU (in our study, each bank is a unit). Using Equation 6, the intertemporal distance functions $D^I(x^{k',s}, y^{k',s}/D^{k',s}(x^{k',s}, y^{k',s})$, $s = t, t+1$ are calculated using the following formula:

$$[D^I(x^{k',s}, y^{k',s}/D^{k',s}(x^{k',s}, y^{k',s})]^{-1} = max \phi_I^{k'}$$

subject to:

$$\sum_{k \in R_j,\, s \in \tau} z^k y_m^k \geq \phi_I^{k'} \hat{\phi}_c^{k',s} y_m^{k',s}, \quad m=1 \ldots \ldots \ldots M$$

$$\sum_{k \in R_j,\, s \in \tau} z^{k,s} x_n^{k,s} \leq x_n^{k',s}, \quad n=1 \ldots \ldots \ldots N$$

$$z^{k,s} \geq 0, \tau = (1,2, \ldots ., T) \quad \text{Equation 7}$$

The above equation examines all units of all time periods for any specific group R_j. Now, the following objective function is responsible for calculating objective function for all units, all periods and all groups in any study. Denoting the solution of Equation 7, the global distance function $D^{Global}(x^{k',s}, y^{k',s}/ D^{k',s}(x^{k',s}, y^{k',s})$, $s = t, t+1$ can be calculated as follows:

$$[D^{Global}(x^{k',s}, y^{k',s}/ D^{k',s}(x^{k',s}, y^{k',s})]^{-1} = max \phi_{Global}^{k'}$$

Subject to:

$$\sum_{k \in R_j,\, s \in \tau} z^k y_m^k \geq \phi_{Global}^{k'} \hat{\phi}_I^{k's} y_m^{k's}, \quad m=1 \ldots \ldots \ldots M$$

$$\sum_{k \in R_j,\, s \in \tau} z^{k,s} x_n^{k's} \leq x_n^{k's}, \quad n=1 \ldots \ldots \ldots N$$

$$z^{k,s} \geq 0, R = R_1 U R_2 U, \ldots ., U R_J, \tau = (1,2, \ldots ., T) \quad \text{Equation 8}$$

Using Equations 6, 7 and 8, the optimal solution for the Malmquist metafrontier index can be calculated and decomposed.

Data and variables

The input and output variables are listed in Table 2. Given the existing heterogeneity among MFIs, we considered the production approach (i.e. traditional factors of production, labour and capital) of MFIs rather than the intermediation approach (i.e. transformation of loans from savings) as proposed by Haq and Skully[4] and Annim[8]. All values are expressed in USD in real rates. We use a balanced panel data of 473 MFIs from five regions over the period 2004–2013. Here, x1 refers to cost per borrower and x2 refers to cost per loan. On the output side, y1 refers to borrowers per staff members, y2 refers to borrowers per loan officer, and y3 refers to depositors per staff members used. Hence, all the input variables are connected to the production approach based on the earlier studies of MFI efficiency.

Table 2: Descriptive statistics of inputs and outputs in microfinance institutions

	Minimum	Maximum	Mean	Standard deviation
Sub-Saharan Africa				
x1: cost per borrower	71	162	99	102
x2: cost per loan	12	321	154	169
y1: borrowers per staff member	17	358	198	217
y2: borrowers per loan officer	88	352	266	281
y3: depositors per staff member	30	451	324	347
Middle East and North Africa				
x1: cost per borrower	18	145	101	124
x2: cost per loan	21	198	166	185
y1: borrowers per staff member	39	354	217	246
y2: borrowers per loan officer	17	274	199	231
y3: depositors per staff member	15.5	297	264	298
East and South Asia				
x1: cost per borrower	9	87	41	17
x2: cost per loan	154	1024	889	354
y1: borrowers per staff member	34	236	163	189
y2: borrowers per loan officer	18.5	400	215	263
y3: depositors per staff member	12	318	264	287
Latin America and the Caribbean				
x1: cost per borrower	31	245	185	134
x2: cost per loan	27	314	226	248
y1: borrowers per staff member	11	187	141	167
y2: borrowers per loan officer	24	287	219	142
y3: depositors per staff member	19.55	301	245	214
Eastern Europe and Central Asia				
x1: cost per borrower	6.5	111	78	19.8
x2: cost per loan	25	206	154	143
y1: borrowers per staff member	16.4	244	169	198
y2: borrowers per loan officer	19	286	217	14.2
y3: depositors per staff member	14.65	182	159	16.8

Empirical results and discussion

Before examining efficiency of MFIs in a metafrontier production function using the proposed model by Oh and Lee[44], it is essential to determine the number of technological groups (here, five selected regions). This will define which MFI belongs to which group. Because the criteria for selecting an MFI for a group was based on our proposed merits (geographical closeness and socio-economic conditions for poverty alleviation) as discussed earlier, an alternative decision and selection choice could lead to different results. Our main objective was to find a group as a leader among the frontiers, that is, technical leadership across the groups.

Table 3 presents productivity changes and their decompositions obtained from Equations 6, 7 and 8. From the results of MI, it is seen that the productivity change among the SSA and MENA countries, on average, declined slightly (less than 1%) over the study period. For the remaining regions, the productivity change scored positive over the study period. The highest productivity progress was observed for ESA, followed by Eastern Europe and Central Asia and then LAC. Among ESA countries, only Pakistan and Cambodia showed mildly negative progress. For Pakistan, our result was in line with the findings of Rauf and Mahmood[49], who showed that cost-per-borrower had increased unreasonably and caused a decrease in efficiency. However, our result for Cambodia contradicted the earlier result of Tahir and Tahrim[17].

Table 3: Mean changes in important indicators

	Productivity change	Efficiency Change	Best practice gap change	TGC change		Productivity change	Efficiency Change	Best practice gap change	TGC change
Sub-Saharan Africa					**Latin America and the Caribbean**				
Cameroon	1.0000	0.9945	0.9852	1.0207	Argentina	1.0013	0.9997	1.0015	1.0012
Congo	1.0012	1.0006	1.0004	1.0003	Bolivia	1.0062	1.0358	0.9813	0.9900
Ghana	1.0077	0.9976	1.0054	1.0047	Brazil	0.9962	1.0088	0.9916	0.9959
Kenya	0.9799	0.9878	0.9909	1.0011	Chile	1.0173	1.0000	1.0046	1.0026
Madagascar	0.9849	0.9774	1.0015	1.0062	Colombia	0.9909	1.0000	0.9785	1.0126
Mozambique	1.0000	0.9909	1.0094	0.9998	Costa Rica	1.0126	1.0104	0.9998	1.0023
Nigeria	1.0013	0.9997	1.0015	1.0012	Dom. Republic	1.0152	1.0229	1.0026	0.9898
Rwanda	0.9897	1.0000	0.9940	0.9957	Ecuador	1.0210	1.0040	0.9907	1.0265
Senegal	1.0032	0.9999	1.0133	0.9902	El Salvador	1.0000	0.9945	0.9852	1.0207
Tanzania	0.9981	0.9900	1.0050	1.0032	Guatemala	1.0012	1.0006	1.0004	1.0003
Uganda	0.9908	0.9967	0.9950	0.9991	Haiti	1.0032	0.9932	1.0054	1.0047
Mean	***0.9961***	***0.9941***	***1.0001***	***1.0020***	Honduras	0.9905	1.0110	0.9901	0.9896
Middle East and North Africa					Jamaica	0.9975	1.0069	0.9926	0.9981
					Mexico	0.9800	0.9972	0.9889	0.9937
Egypt	0.9967	0.9937	1.0017	1.0013	Nicaragua	1.0021	0.9919	1.0048	1.0055
Lebanon	0.9967	0.9937	1.0017	1.0013	Panama	1.0062	1.0080	1.0022	0.9960
Morocco	1.0023	1.0054	0.9983	0.9986	Paraguay	0.9958	1.0104	0.9912	0.9942
Yemen	0.9950	1.0016	0.9944	0.9990	Peru	1.0059	0.9967	1.0052	1.0040
Mean	***0.9977***	***0.9986***	***0.9990***	***1.0001***	Trinidad & Tobago	0.9790	0.9876	0.9939	0.9973
East and South Asia					***Mean***	***1.0006***	***1.0042***	***0.9953***	***1.0013***
					Eastern Europe and Central Asia				
Cambodia	0.9881	0.9753	1.0129	1.0002					
China	1.0326	1.0104	0.9998	1.0023	Armenia	1.0048	0.9994	1.0066	1.0017
Indonesia	1.0152	1.0229	1.0026	0.9898	Azerbaijan	1.0038	1.0165	0.9945	0.9989
Philippines	1.0010	1.0040	0.9907	1.0265	Bosnia and Herzegovina	1.0057	1.0119	1.0020	1.0018
Vietnam	1.0000	0.9945	0.9852	1.0207	Georgia	1.0124	1.0010	1.0088	1.0026
Bangladesh	1.0062	1.0080	1.0022	0.9960	Kazakhstan	0.9885	0.9843	1.0004	1.0039
India	1.0077	0.9976	1.0054	1.0047	Kyrgyzstan	0.9975	1.0069	0.9926	0.9981
Pakistan	0.9799	0.9878	0.9909	1.0011	Tajikistan	0.9800	0.9972	0.9889	0.9937
Mongolia	1.0143	1.0216	0.9905	1.0023	Turkey	1.0021	0.9919	1.0048	1.0055
Mean	***1.0050***	***1.0025***	***0.9978***	***1.0048***	***Mean***	***1.0017***	***1.0011***	***0.9998***	***1.0008***

According to their findings, Cambodian MFI productivity was increasing because of consistent success of scale efficiency, which implies that MFIs were operating at maximum scales. The highest productivity change in the ESA region was noted for China (3.2% annually). This finding confirms earlier studies.[50] According to those earlier studies, strong macroeconomic support helped China to achieve good progress in productivity changes over the years.

In terms of efficiency changes, again SSA and MENA countries were on average shown to have dropped by less than 1% over the years. The greatest progress among these countries was observed for Morocco. Efficiency change reportedly progressed annually for the remaining three regions, with the greatest progress being observed for LAC. The latter finding again contradicted the results of an earlier study by Servin and Lensink[41], who reported that in general MFIs in LAC showed a drop in efficiency scores. Among all the countries, the highest annual efficiency decrease was no more than 1.5%. This signifies that MFIs, worldwide, have been operating at almost close to unit efficiency on average. This particular result highlights that most MFIs across these countries are not lagging far behind the group frontier.

The measure of technical change or business process change indicates whether countries have technically progressed or not. It is evident from Table 3 that on average, all regions except SSA showed a drop in efficiency in technical change. This means that African countries, on average, are technically progressing over time although the progress is relatively very small. This result is in line with an earlier study by Lafourcade and Isern[14]. A further interesting finding with regard to business process change is that the remaining four regions are lagging behind in unit efficiency by less than 1%. Thus, technical efficiencies among the selected countries do not deviate significantly from each other.

The measure of how much a country gets closer to or further away from world frontier technology is known as the TGC score. The average TGC score is increasing yearly at 0.48% for ESA countries, which shows that they are approaching the frontier. An interesting result is that every ESA country has TGC score of 1 and most SSA countries have TGC scores less than 1. This signifies that ESA countries are leading the world frontier technology and most African countries are not.

Conclusion

Our study examined the issue of heterogeneity among MFIs worldwide.[20] We examined production efficiency of MFIs in Africa and their capacity for poverty alleviation. Out of 929 MFIs listed by the Economist Intelligence Unit, our study included 743 (79.98%) MFIs in 51 countries. Based on regional closeness and socio-economic similarities, we classified these countries into five regional groups. Because our study examined production efficiency, only MFIs that offered both saving and lending facilities were included in the final selection. More than 90% of the MFIs that existed in ESA and LAC at the time of the study were included in our sample. The sample proportions for the regions of SSA, MENA, and Eastern Europe and Central Asia were 60%, 57% and 58% respectively. The total data set included 7430 observations from the 743 MFIs worldwide, covering the years 2004 to 2013.

To examine efficiency, we used two input variables and three output variables. Methodologically, we used metafrontier technology to examine the heterogeneity of MFIs. Initially, the metafrontier MI was used to calculate the technical efficiency, technology gap for the best practice and technological gap ratio among the studied banks. For the efficiency calculation, we ran non-parametric DEA.

The empirical results from our study show that regionally there is a link between group performance and individual performance, which reflects the findings of earlier studies of heterogeneity.[10,20] For instance, when identifying technological growth change among the regions, we noted that the highest TGC score occurred in ESA countries. The analysis also showed that all ESA countries had a TGC score of more than 1. A comparison of EC, SSA and MENA countries showed that on average their TGC scores had dropped by less than 1% over the years. The greatest progress among these countries was observed for Morocco. By region, the greatest progress was observed for LAC countries.

The main contribution of this paper is it successfully blends two streams of academic theory, namely the MFI institutional paradigm and theories of efficiency. The results reveal that social outreach is a sequential event that arises from efficient performance of MFIs in a society. Financially sound MFIs can contribute to society by taking higher risks and being flexible about the terms of repayment.

We also introduced MFI heterogeneity as the key measure to compare MFIs' performances worldwide. Thus, benchmarking becomes robust and policy-making for future planning becomes more resourceful. Despite MFIs in some regions having greater financial efficiency through the scale of economy (scale efficiency), they might suffer from deficits in management skills. However, such a result contradicts the earlier studies[29]. Thus, we identified a gap in examining how MFI efficiency based on managerial skill may further improve MFIs' performance worldwide.

The findings from this study have important implications for managers and policymakers. From the regulatory perspective, MFIs can now be grouped, which means that benchmarking can be based on the group that an MFI belongs to. Internationally, managers can explain other MFIs' performance for benchmarking and decision-making. Heterogeneity is a major concern for benchmarking efficiency. The diverse nature of business and the complexity in defining variables make the efficiency calculation even more difficult. Unidentified or unobserved variables might have a significant influence on banks' efficiency. Nonetheless, our empirical results show that geographical location, strategic and socio-economic conditions, and macroeconomic heterogeneity can significantly influence MFI efficiency.

Authors' contributions

S.M. was the project leader; M.A.K.A. and A.K.M.M. were responsible for data analysis and project design; M.A.K.A., A.K.M.M. and P.W. performed the data analysis; M.A.K.A., A.K.M.M., S.M. and P.W. made conceptual contributions; and S.M. and P.W. performed the analyses. M.A.K.A. and A.K.M.M. wrote the manuscript and S.M. edited and approved it.

References

1. Hermes N, Lensink R, Meesters A. Outreach and efficiency of microfinance institutions. World Dev. 2011;39(6):938–948. http://dx.doi.org/10.1016/j.worlddev.2009.10.018
2. Woller GM, Dunford C, Woodworth W. Where to microfinance. Int J Econ Dev. 1999;1(29):29–64.
3. Morduch J. The microfinance schism. World Dev. 2000;28(4):617–629. http://dx.doi.org/10.1016/S0305-750X(99)00151-5
4. Haq M, Skully M, Pathan S. Efficiency of microfinance institutions: A data envelopment analysis. Asia-Pac Finance Markets. 2010;17(1):63–97. http://dx.doi.org/10.1007/s10690-009-9103-7
5. Sengupta R, Aubuchon CP. The microfinance revolution: An overview. St Louis Rev: Fed Reserve Bank; 2008. pages 9–30.
6. Gutierrez-Nieto B, Serrano-Cinca C, Molinero CM. Social efficiency in microfinance institutions. J Oper Res Soc. 2009;60(1):104–119. http://dx.doi.org/10.1057/palgrave.jors.2602527
7. Sainz-Fernandez I, Torre-Olmo B, López-Gutiérrez C, Sanfilippo-Azofra S. Crisis in microfinance institutions: Identifying problems. J Int Dev. 2015;27(7):1058–1073. http://dx.doi.org/10.1002/jid.3129
8. Annim SK. Microfinance efficiency: Trade-offs and complementarities between the objectives of microfinance institutions and their performance perspectives. Europ J Devel Res. 2012;24(5):788–807. http://dx.doi.org/10.1057/Ejdr.2011.60
9. Wagner C, Winkler A. The vulnerability of microfinance to financial turmoil: Evidence from the global financial crisis. World Dev. 2013;51:71–90. http://dx.doi.org/10.1016/j.worlddev.2013.05.008
10. Kareithi RN, Lund C. Review of NGO performance research published in academic journals between 1996 and 2008. S Afr J Sci. 2012;108(11–12):36–44. http://dx.doi.org/10.4102/sajs.v108i11/12.755

11. Van Rooyen C, Stewart R, De Wet T. The impact of microfinance in sub-Saharan Africa: A systematic review of the evidence. World Dev. 2012;40(11):2249–2262. http://dx.doi.org/10.1016/j.worlddev.2012.03.012

12. Hulme D, Mosley P. Finance against poverty. New York, USA: Psychology Press; 1996.

13. Buckley G. Microfinance in Africa: Is it either the problem or the solution? World Dev. 1997;25(7):1081–1093. http://dx.doi.org/10.1016/S0305-750X(97)00022-3

14. Lafourcade A-L, Isern J, Mwangi P, Brown M. Overview of the outreach and financial performance of microfinance institutions in Africa. Washington, DC: Microfinance Information Exchange; 2005.

15. Bassem BS. Efficiency of microfinance institutions in the Mediterranean: An application of DEA. Transit Stud Rev. 2008;15(2):343–354. http://dx.doi.org/10.1007/s11300-008-0012-7

16. Bassem BS. Total factor productivity change of MENA microfinance institutions: A Malmquist productivity index approach. Econ Model. 2014;39:182–189. http://dx.doi.org/10.1016/j.econmod.2014.02.035

17. Tahir IM, Tahrim SNC. Efficiency and productivity analysis of microfinance institutions in Cambodia: A DEA approach. Int Rev Bus Res Pap. 2015;11(1):25–42.

18. Habib M, Jubb C. Poverty alleviation with microfinance: Bangladesh evidence. In Heshmati A, Maasoumi E, Wan G, editors. Poverty reduction policies and practices in developing Asia. Economic studies in inequality, social exclusion and well-being. Singapore: Springer; 2015. p. 11–30. http://dx.doi.org/10.1007/978-981-287-420-7_2

19. Azad MAK, Masum A, Munisamy S, Sharmin D. Efficiency analysis of major microfinance institutions in Bangladesh: A Malmquist index approach. Qual Quant. 2016;50(4):1525–1537. http://dx.doi.org/10.1007/s11135-015-0219-8

20. Bolli T, Vo Thi A. Regional differences in the production processes of financial and social outputs of microfinance institutions. Econ Transition. 2014;22(3):461–495. http://dx.doi.org/10.1111/ecot.12045

21. Guntz S. Sustainability and profitability of microfinance institutions. App Int Fin Dev. 2011;4:1–53.

22. 22. Louis P, Seret A, Baesens B. Financial efficiency and social impact of microfinance institutions using self-organizing maps. World Dev. 2013;46:197–210. http://dx.doi.org/10.1016/j.worlddev.2013.02.006

23. Hayami Y. Sources of agricultural productivity gap among selected countries. Amer J Agr Econ. 1969;51(3):564–575. http://dx.doi.org/10.2307/1237909

24. Binswanger HP, Ruttan VW. Induced innovation: Technology, institutions, and development. Baltimore: Johns Hopkins University Press; 1978.

25. Battese GE, Rao DP. Technology gap, efficiency, and a stochastic metafrontier function. Int J Bus Econ. 2002;1(2):87–93

26. Battese GE, Rao DP, O'Donnell CJ. A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. J Productiv Anal. 2004;21(1):91–103. http://dx.doi.org/10.1023/B:PROD.0000012454.06094.29

27. Liu JS, Lu LYY, Lu W-M, Lin BJY. A survey of DEA applications. Omega-Int J Manage S. 2013;41(5):893–902. http://dx.doi.org/10.1016/j.omega.2012.11.004

28. World Bank. Poverty Overview 2015 [document on the Internet]. 2015 [cited 2015 May 5]. Available from: http://www.worldbank.org/en/topic/poverty/overview.

29. Im J, Sun S. Profits and outreach to the poor: The institutional logics of microfinance institutions. Asia Pac J Manag. 2015;32(1):95–117. http://dx.doi.org/10.1007/s10490-014-9398-4

30. Yunus M. Creating a world without poverty: Social business and the future of capitalism. New York: Public Affairs; 2007.

31. Farrington T. Efficiency in microfinance institutions. Microbanking Bulletin. 2000;4:18–23.

32. Farrell MJ. The measurement of productive efficiency. J Roy Stat Soc A Sta. 1957;120(30):253–290. http://dx.doi.org/10.2307/2343100

33. Berger AN, Humphrey DB. Efficiency of financial institutions: International survey and directions for future research. Eur J Oper Res. 1997;98(2):175–212. http://dx.doi.org/10.1016/S0377-2217(96)00342-6

34. Lampe HW, Hilgers D. Trajectories of efficiency measurement: A bibliometric analysis of DEA and SFA. Eur J Oper Res. 2015;240(1):1–21. http://dx.doi.org/10.1016/j.ejor.2014.04.041

35. Charnes A, Cooper W, Rhodes E. Measuring the efficiency of decision making units. Eur J Oper Res. 1978;2(6):429–444. http://dx.doi.org/10.1016/0377-2217(78)90138-8

36. Charnes A, Cooper W, Lewin AY, Seiford LM. Data envelopment analysis theory, methodology and applications. J Oper Res Soc. 1997;48(3):332–333. http://dx.doi.org/10.1057/palgrave.jors.2600342

37. Cooper W, Seiford L, Tone K. Introduction to data envelopment analysis and its uses: With DEA-solver software and references. US: Springer Science & Business Media; 2006.

38. Agarwal PK, Sinha SK. Financial performance of microfinance institutions of India. Delhi Bus Rev. 2010;11(2):37–46.

39. Nghiem H, Coelli T, Rao P. The efficiency of microfinance in Vietnam: Evidence from NGO schemes in the north and the central regions. Int J Environ Cul Econ Soc Sus. 2006;2(5):71–78.

40. Wijesiri M, Viganò L, Meoli M. Efficiency of microfinance institutions in Sri Lanka: A two-stage double bootstrap DEA approach. Econ Model. 2015;47:74–83. http://dx.doi.org/10.1016/j.econmod.2015.02.016

41. Servin R, Lensink R, van den Berg M. Ownership and technical efficiency of microfinance institutions: Empirical evidence from Latin America. J Bank Financ. 2012;36(7):2136–2144. http://dx.doi.org/10.1016/j.jbankfin.2012.03.018

42. Marconatto D, Barin Cruz L, Avila Pedrozo E. Going beyond microfinance fuzziness. J Clean Prod. 2016;115:5–22. http://dx.doi.org/10.1016/j.jclepro.2015.12.070

43. Allet M. Why do microfinance institutions go green? An exploratory study. J Bus Ethics. 2014;122(3):405–424. http://dx.doi.org/10.1007/s10551-013-1767-2

44. Oh D-h, Lee J-d. A metafrontier approach for measuring Malmquist productivity index. Empir Econ. 2010;38(1):47–64. http://dx.doi.org/10.1007/s00181-009-0255-0

45. Pastor JT, Lovell CAK. A global Malmquist productivity index. Econ Lett. 2005;88(2):266–271. http://dx.doi.org/10.1016/j.econlet.2005.02.013

46. Tulkens H, Vanden Eeckaut P. Non-frontier measures of efficiency, progress and regress for time series data. Int J Product Econ. 1995;39(1–2):83–97. http://dx.doi.org/10.1016/0925-5273(95)00002-6

47. Caves D, Christensen L, Diewert W. The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica. 1982;6:1393–1414. http://dx.doi.org/10.2307/1913388

48. Fare R, Grosskopf S, Norris M, Zhang Z. Productivity growth, technical progress and efficiency changes in industrial countries. Amer Econ Rev. 1994;84:66–83.

49. Rauf SA, Mahmood T. Growth and performance of microfinance in Pakistan. Pakistan Econ Soc Rev. 2009;74(1):99–122.

50. Cheng E, Ahmed AD. The demand for credit, credit rationing and the role of microfinance. China Agr Econ Rev. 2014;6(2):295–315. http://dx.doi.org/10.1108/CAER-07-2012-0076

San and Nama indigenous knowledge: The case of |*nhora* (*Pteronia camphorata*) and its medicinal use

AUTHORS:
Isabel M. Hulley[1]
Patricia M. Tilney[1]
Sandy F. Van Vuuren[2]
Guy P.P. Kamatou[3]
Janneke M. Nortje[1]
Alvaro M. Viljoen[3]
Ben-Erik Van Wyk[1]

AFFILIATIONS:
[1]Department of Botany and Plant Biotechnology, University of Johannesburg, Johannesburg, South Africa
[2]Department of Pharmacy and Pharmacology, University of the Witwatersrand, Johannesburg, South Africa
[3]Department of Pharmaceutical Sciences, Tshwane University of Technology, Pretoria, South Africa

CORRESPONDENCE TO:
Ben-Erik Van Wyk

EMAIL:
bevanwyk@uj.ac.za

DATES:
Received: 04 Feb. 2016
Revised: 05 May 2016
Accepted: 06 May 2016

KEYWORDS:
Asteraceae; essential oil; ethnobotany; leaf anatomy; antimicrobial activity

HOW TO CITE:
Hulley IM, Tilney PM, Van Vuuren SF, Kamatou GPP, Nortje JM, Viljoen AM, et al. San and Nama indigenous knowledge: The case of |*nhora* (*Pteronia camphorata*) and its medicinal use. S Afr J Sci. 2016;112(9/10), Art. #2016-0044, 9 pages. http://dx.doi.org/10.17159/sajs.2016/20160044

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa)

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

A hitherto unidentified medicinal plant is here identified for the first time as *Pteronia camphorata* (L.) L., an aromatic shrub of the Asteraceae family endemic to the western and southern coastal region of South Africa. The plant was described in this journal by Laidler[1] in 1928 as '*D/nhora* buchu', and is one of the important types of buchu used by the Nama people. We report the traditional medicinal uses among San and Nama people, based on our interviews with rural participants. These include the treatment of colds, influenza, chest ailments and tuberculosis, as well as convulsions, haemorrhoids and inflammation of the neck. The major and minor chemical compounds of the essential oil that is produced by the plant are identified, together with the site of accumulation of this volatile oil within the leaf. We also investigated the plant's antimicrobial activity against a selection of a yeast and two Gram-negative and one Gram-positive bacteria, all of which are associated with respiratory infections. *P. camphorata* is scientifically poorly known but is an important San and Nama traditional remedy. Our study not only prevents the potential loss of historically important indigenous knowledge, but also provides the first scientific evidence to validate the traditional use of |*nhora* against upper and lower respiratory tract infections, including tuberculosis. This detailed study has wider application in demonstrating the fragility of the oral–traditional knowledge of a scientifically neglected indigenous group. It also highlights the scientific and practical importance of preserving traditional plant-use knowledge within a botanically diverse region.

Significance:

- Reveals the botanical identity of |*nhora*, an important Nama medicinal plant.
- Presents scientific evidence to validate the traditional uses.
- Contributes to the cultural heritage of a scientifically neglected indigenous group.
- Demonstrates the fragility of oral–traditional knowledge.

Introduction

The traditional Khoi and San practices of plant use in the botanically rich Cape region of South Africa are poorly documented. They are also rapidly diminishing because of the fragility of orally transferred indigenous knowledge, and the fact that modern medicine has almost completely replaced the traditional health care system. Because the Khoi and San peoples are ancestral to the rest of humanity, their ethnobotanical knowledge is of global significance. It provides insights into the early history and origins of medicinal plant use. We therefore hope to prevent the loss of profound ethnobotanical information relating to an important fynbos-endemic medicinal plant. Fynbos is a natural shrubland or heathland vegetation located in the south-western part of South Africa.

The hitherto unidentified plant described by Laidler[1] in 1928 as '*D/nhora'*, one of the important types of buchu used by the Nama people, is here identified for the first time as *Pteronia camphorata* (L.) L. (Asteraceae). 'Buchu' (*boegoe*) refers to aromatic plants that are used, often in powdered form, for medicinal and cosmetic purposes. Laidler used *D/* to denote the voiceless (tenuis) dental click, but we use the pipe symbol (|) in accordance with the International Phonetic Alphabet. Some linguists still prefer the (ǀ) symbol of the Doke / Beach convention.

The cultural and historical importance of this species came to light gradually, mainly because of ethnobotanical field survey work in Namaqualand.[2] Only ten elderly persons, two of whom are now deceased, had traditional knowledge about *P. camphorata* and its use in San and Nama medicine. *Pteronia* L. is a genus of roughly 70 small shrubs subendemic to southern Africa[3] and has been the focus of recent ethnobotanical studies[4–6]. Although these aromatic plants are of considerable local importance as sources of traditional medicine, their ethnobotany, leaf anatomy, essential oil chemistry and antimicrobial activity had previously not been studied systematically. *P. camphorata* is a small shrub of up to 0.8 m in height; it has a wide distribution in the Northern, Western and Eastern Cape provinces of South Africa. Four varieties have been distinguished (Figure 1), based on leaf arrangement (alternate or opposite) and the presence and density of white hairs (cilia) on leaf surfaces and margins.[3]

We wish to dispel the myth that indigenous knowledge is mainly of cultural and symbolic value, with limited practical use in health care. Our study explores the scientific rationale for the main use of |*nhora* in treating respiratory infections, including tuberculosis. For this purpose we conducted a detailed scientific study of the ethnobotany, leaf anatomy, essential oil chemistry and antimicrobial activity of *P. camphorata.*

Materials and methods

Materials studied

Three fresh samples of *P. camphorata* var. *armata* were collected at Kleinvlei, close to Wupperthal (32° S, 19° E), for the purpose of essential oil analysis and antimicrobial studies. Samples of all four varieties of *P. camphorata* were collected at various localities for anatomical study. Exact localities, voucher specimen details and authorities for names are shown in Table 1.

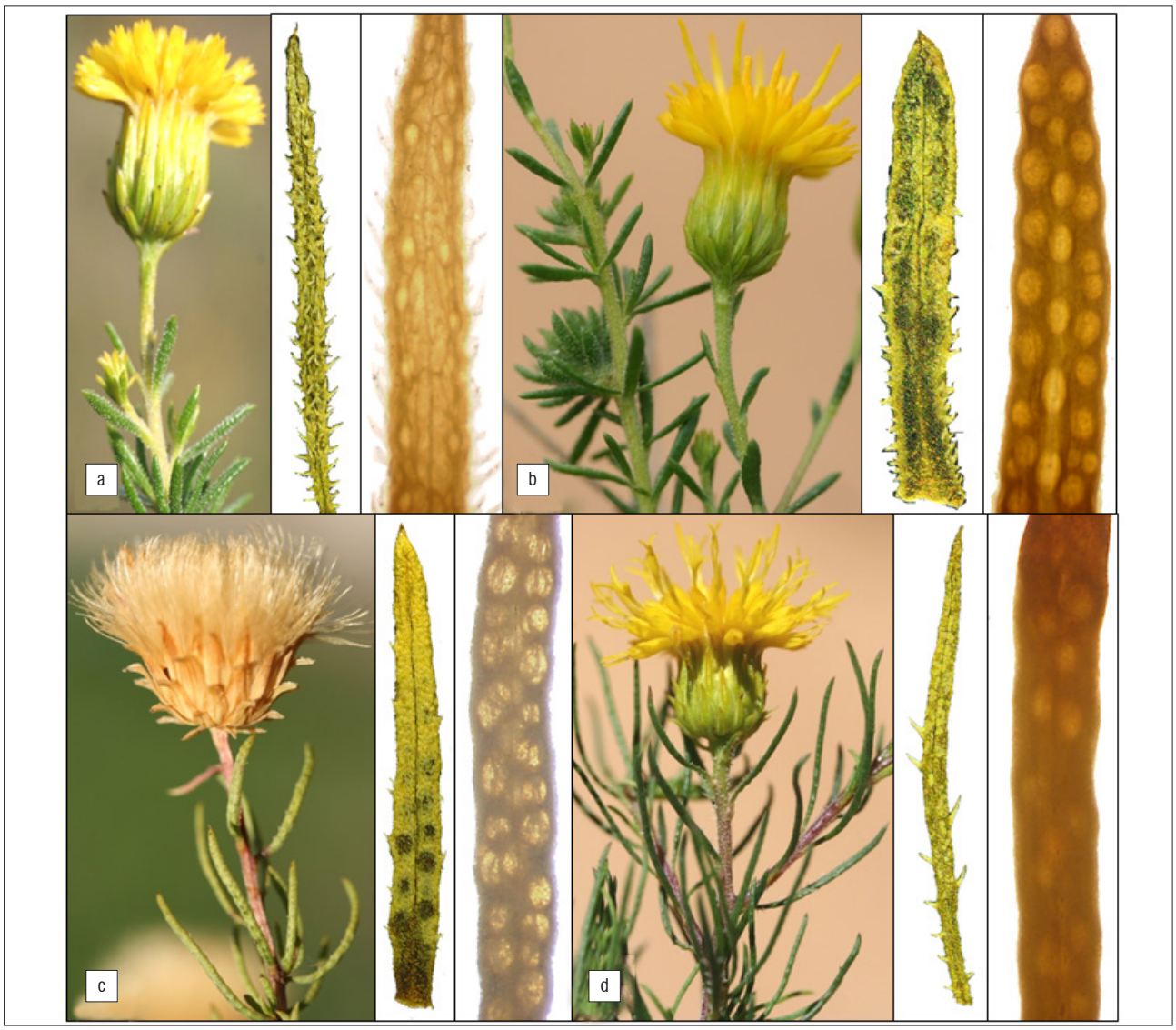


Figure 1: The morphology of flower-heads and leaves of four varieties of *Pteronia camphorata*. Specimens were (a) *P. camphorata* var. *armata* at Kleinvlei near Wupperthal, (b) *P. camphorata* var. *camphorata* at Dasklip Pass near Porterville, (c) *P. camphorata* var. *laevigata* at Nieuwoudtville, and (d) *P. camphorata* var. *longifolia* from Paarl Mountain. Note the distribution of cilia on leaves and presence of translucent oil glands. Photographs by B.E. Van Wyk.

Table 1: Voucher specimen details of the plant materials of *Pteronia camphorata* that were studied

Variety	Locality	Date collected	Voucher specimens (all housed in JRAU)	Anatomy (A) Extracts for antimicrobial (MIC) and chemical composition (GC–MS)
P. camphorata (L.) L. var. *armata* Harv.	Kleinvlei, Wupperthal	4 October 2009	*B.-E. Van Wyk, I.M. Hulley & P.M. Tilney 4443*	MIC, GC–MS, A
P. camphorata (L.) L. var. *camphorata*	Dasklip Pass, Porterville	3 October 2009	*B.-E. Van Wyk, I.M. Hulley & P.M. Tilney 4428*	A
P. camphorata (L.) L. var. *laevigata* Harv.	Nieuwoudtville, Oorlogskloof	19 August 1997	*B.-E. Van Wyk 3702*	A
P. camphorata (L.) L. var. *laevigata* Harv.	Nieuwoudtville, Oorlogskloof	11 December 2008	*B.-E. & M. Van Wyk 4287*	A
P. camphorata (L.) L. var. *longifolia* Harv.	Paarl Mountain	3 October 2009	*B.-E. Van Wyk, I.M. Hulley & P.M. Tilney 4427*	A

Ethnobotanical interviews

The recording of ethnobotanical data strictly adhered to all the ethical principles in the Code of Ethics[7] of the International Society of Ethnobiology. These include the concepts of traditional resource rights, educated prior informed consent, and respect for the privacy and customs of rural people. Formal approval was also obtained (by B.E.V.W. and J.M.N.) from the ethics committee of the Faculty of Science at the University of Johannesburg. The data presented here form part of two extensive ethnobotanical surveys in Namaqualand and the Cederberg region. The ethical clearance reference numbers are UJ Protocol No. 13 January 2015 (B.E.V.W., Cederberg survey and other studies); UJ Protocol No. 20 May 2011 (J.M.N., Kamiesberg survey, for MSc) and UJ Protocol No. 7 July 2015 (J.M.N., Namaqualand survey, for PhD).

Interviews were conducted in the local language, Afrikaans, and we used the matrix method as previously described[2]. An extensive network of local participants was interviewed: close to 100 people in Namaqualand and more than 30 people in the Cederberg region, including several local experts (Table 2). Our paper provides a firm starting date and geographical localities for the traditional knowledge associated with *P. camphorata*, including the identities of the knowledge holders.

Anatomical procedures

Dried leaf material of *P. camphorata* var. *armata* was rehydrated and placed in formaldehyde-acetic acid–alcohol (formaldehyde [40%] : glacial acetic acid : ethanol [96%] : distilled water, 2:1:10:7) together with fresh leaf material from the remaining three varieties, for 24 h. Thereafter the material was treated according to the methods described previously[4–6] (dehydration, infiltration with and embedding in glycol methacrylate, sectioning, staining and data capturing). Unstained slides were viewed under a polarising microscope to observe whether any crystals were present.

Distillation and analysis of essential oil

Leaves and twigs (roughly 300 g dry weight) from three different plants of *P. camphorata* var. *armata* (Table 3) were air-dried for 9 days, then subjected to hydro-distillation for 180 min using a Clevenger-type apparatus. The oils were weighed and stored in sealed vials in the dark at 4 °C before analysis.

The oils (20% diluted in hexane) were analysed by a gas chromatography–mass spectrometry system using standardised settings, as described previously[4]. Compounds were identified by their mass spectral data and retention indices, as well as library searches of the NIST®, Mass Finder® and Flavour® libraries.

Antimicrobial studies

Various extracts (Table 4) and the oil samples listed in Table 3 were investigated for antimicrobial activities, using the minimum inhibitory concentration (MIC) microtitre plate method as previously described[4]. A yeast (*Cryptococcus neoformans*, American type culture collection [ATCC] 90112) as well as two Gram-negative bacteria (*Moraxella catarrhalis* ATCC 23246 and *Klebsiella pneumoniae* ATCC 13883) and one Gram-positive bacterium (*Mycobacterium smegmatis* ATCC 14468) were selected for the study. All cultures were selected on the basis of their respiratory pathogenesis, with the exception of *M. smegmatis*, which is a non-pathogenic *Mycobacterium* strain with a faster growth rate than that of *M. tuberculosis* – hence its ease of use. *Cryptococcus neoformans* is associated with lung infection, and *M. catarrhalis* is known to cause bronchitis, sinusitis and laryngitis. *Klebsiella pneumoniae* commonly causes pneumonia. The preparation of bacterial cultures and the methodology followed, including the preparation of extracts, were exactly the same as described previously.[4–6] MIC assays were undertaken in duplicate or triplicate on separate occasions, and the mean results are presented.

Results and discussion

Ethnobotany

A historically important ethnobotanical paper by Laidler[1], published in this journal in 1928, gave the names of several species of buchu used in Namaqualand. The use of buchu (powdered aromatic bushes, usually stored in tortoise shells) as a topical treatment of the skin is of San origin, and is well described in the literature.[4,8] Buchu, sab or *P/nkaou* (Laidler used *P/* to denote the palatal or palato-alveolar click, ǂ) is said to have two varieties. The first is *D/nhora* (|*nhora*), *D/khonsa* (|*khonsa*) or *haas buchu*, and the other is *P/kabourie* (ǂ*kabourie*). The botanical identities of these two plants have remained unknown, perhaps because Laidler was not a botanist or he saw only the powdered plant material. He described the use of the two buchu types as follows[1]:

> *Considered by Hottentot to possess great virtues in curing disorders, and rarer forms are valuable to the Native. In powder used for dusting, for fire and sunburns Baby powder made of D/nhora kept in tortoise shell puff box, roasted with fat and dropped in ear for earache. Powder also mixed with C/ghoonabi, and acacia thorn sucker, and then considered a high class powder. The second rate powder used for babies was mixed with ground acacia thorn bark, and is named O/kai. D/nhora Buku is roasted with oil and fat, mixed with mother's milk and dropped into the ear for earache. P/nkaou is rare and valuable, and a thimbleful is worth from a sheep to an ox. It is obtained in mountainous districts and when used is pounded and tied in a cloth over which hot water is poured.*

According to Laidler[1], the name *P/nkaou* was given to him for what was known as 'Buchu barosma' or 'Letulina' at the time. This clearly refers to round-leaf buchu, *Agathosma betulina* (P.J. Bergius) Pillans, a species confined to the mountainous region of the Bokkeveld, Cederberg and Grootwinterhoek. It is understandable that this plant would be considered rare and valuable, because it was not available locally but had to be transported over a considerable distance from the Cederberg to Namaqualand; the Kamiesberg is roughly 300 km north of the Cederberg. *P/kabourie* (ǂ*kabourie*) is not mentioned again, and the reader is left with the impression that this name also refers to *P/nkaou*.

The identification of *D/nhora* (|*nhora*) as *P. camphorata* is the result of a few anecdotes spread over more than a decade. The first and most convincing was the explicit information given by Mr Willem 'Blikkies' Steenkamp at Nieuwoudtville, an area where the Namaqualand variety of the species (var. *armata*) also occurs naturally. All available ethnobotanical information on *P. camphorata* is shown in Table 2. Original anecdotes, recorded in Afrikaans, have been retained to ensure that subtle nuances about the plant and its uses are not lost in translation. *P. camphorata* shares with several other species the Afrikaans name *gombossie*.[9–11]

The first reference to *P. camphorata* as a medicinal plant appears on a herbarium specimen (*E.B. Watermeyer 6350* in PRE). According to the label information, the specimen was collected in January 1925 and the plant is known as *koortsbos* (meaning *koorsbos*, i.e. fever bush). Unfortunately, the locality is given imprecisely as 'Little Namaqualand'. The first-ever published ethnobotanical use for *P. camphorata*, namely the treatment of boils, was reported by Watt and Breyer-Brandwijk[12], who cite Kling[13] as the source of their information. Curiously, the species was not included in Rev. Kling's[13] booklet of 1923, so the origin of the information remains a mystery and appears to be based on a misinterpretation – uncritically cited by Arnold et al.[14]

The only other published ethnobotanical anecdote for the species was reported by Van Wyk and Gericke[15], who cited Willem 'Blikkies' Steenkamp, an elderly man of Khoisan descent from the farm Oorlogskloof near Nieuwoudtville in the Northern Cape province of South Africa. The origin of the knowledge is Mr Steenkamp's grandfather, who was ethnically pure San. The plant is regarded as the most important of all the traditional medicines in the area, and is known locally as *norraboegoe*.

Table 2: Summary of vernacular names and ethnobotanical anecdotes for *Pteronia camphorata*

Vernacular names	Anecdotes and their origin
koortsbos: Herbarium specimen label (January 1925, *E.B. Watermeyer 6350* [PRE])	–
–	Watt and Breyer-Brandwijk, 1962[12]: 'A plaster, made from *P. camphorata* L., is used in the Western Cape for drawing boils'. This is an error and cites Kling, 1923[13] as the source.
gombossie: Marloth, 1932[9]; Smith, 1966[10]	–
gombos, ghombossie: Burman and Bean, 1985[11]	–
norraboegoe: Van Wyk and Gericke, 2000[15], citing Willem 'Blikkies' Steenkamp	*'Getrek in kookwater vir verkoue en griep; vir grootmense, gebruik soos dit is (baie bitter); vir kinders, gooi suiker by en kook tot dit stroperig is.'* [Infused in boiling water for colds and influenza; for adults, use as is (very bitter); for children, add sugar and boil until syrupy.]
t/gôrroboegoe: Anna Brand (Nourivier, Kamiesberg)	–
t/ôrro: Elizabeth Kardinaal (Leliefontein, Kamiesberg)	*'Hy is 'n boegoe, gebruik vir kinders met steek in die oor saam met moedersmelk, gooi in die oor en maak toe met 'n pluisie (watte); word gemeng met melk ook vir grootmense met steek in die oor.'* [It is a buchu, used as medicine for earache for infants, mixed with mother's milk as ear drops (cover with cotton wool); mixed with milk and used by adults to treat ear infection.]
koorsbos: Gertruida Brand (Paulshoek, Kamiesberg)	*'Trek hom af vir rumatiek.'* [An infusion used for rheumatism.]
t/ôrroboegoe: Jakobus Brand (Nourivier, Kamiesberg)	*'As poeier gebruik vir kinders se steek in die oor.'* [Used as powder for earache in children.]
t/ôrro: Jan van der Westhuizen (Garies, originally from Paulshoek but also lived in Leliefontein)	*'Die blaar word poeier gemaak en dan tee, dit is bitter, en gebruik vir winde.'* [Powdered leaf used as bitter infusion for flatulence.]
t/ôrro: Pieter Dirkse (Paulshoek, Kamiesberg)	–
wakkerbos: Sors Cloete (Paulshoek, Kamiesberg)	–
t/nouroeboegoe, t/nôrraboegoe, t/nôrro: Anna Stewe (Leliefontein, Kamiesberg)	(1) '*Die plant het smal blaartjies. Gebruik as medisyne, snuif, saam met ander kruie; vir tandpyn (trek die takkie in melk of water); vir steek in die oor (vat die droë blare en maak `n poeier wat op watte in die oor gesit word) – dit het `n pyndodende effek; trek ook daai wind uit.'* [The plant has narrow leaves. Use as medicine, snuff, with other herbs; for toothache (steep a twig in milk or water); for pain in the ear (take the dry leaves and make a powder that is placed on cotton wool in the ear) – it has an analgesic effect; it also 'pulls out that wind' (relieves flatulence).] (2) '*Org word gebruik vir TB. Dit word saam met jantjie-bêrend en t/nouroeboegoe gebruik. Kookwater word op die blare gegooi en laat trek, soos `n tee. Drink drie maal per dag so 'n kelkie vol. Dit is die beste medisyne vir TB. `n Ou vrou het dit gebruik, as die TB pasiënte huis toe gestuur word, maar hulle is nog nie heeltemal gesond nie. Dit het hulle gehelp om aan te sterk'.* [*Org* (*Notobubon pearsonii*) is used as medicine for TB. It is used in combination with *jantjie-bêrend* (*Sutherlandia frutescens*) and *t/nouroeboegoe* (*Pteronia camphorata*). Boiling water is poured on the leaves and allowed to steep, like a tea. Drink a small glassful three times per day. It is the best medicine for TB. An old lady used it when TB patients were sent home before they were fully recovered. It helped them to regain their strength.]
t/nôrro, t/nôrra, t/gôrrôboegoe: Gert Dirkse (Paulshoek, Kamiesberg)	'*Dis `n bos. Die wortel maal dit fyn, vir babatjies, stuipe en vir grootmense. Gebruik vir vuil bloed en siek voel, ook vir grootmens stuipe. Die blare word droog en fyn gebruik. Vir siek voel, meng met baarbos en slangbos. Dit word gekook. Baie goed*'. [It is a bush. The root, grind it to a powder, for babies, convulsions and for adults. Use for impure blood and when feeling ill, also for adult convulsions. The leaves are used dried and powdered. For 'feeling ill', mix with *baarbos* (*Limeum africanum*) and *slangbos* (*Stoebe plumosa*). It is boiled (a decoction). Very good.]
t`ôrrôboegoe* or *t`ôrrôbosboegoe: Sarah Fortuin (Spoegrivier, Namaqualand)	'*t`ôrrôboegoe of t`ôrrôbosboegoe – die t`ôrrô verwys na die nek; dit word gebruik veral vir kinders vir ontsteking om en agter in die nek (as dit so rooi word). Dit is ook `n bors medisyne; hierdie bos is baie goeie medisyne*.' [The *t`ôrrô* refers to the back of the neck; it is used especially for children, to treat inflammation around and behind the neck (when it turns red). It is also a chest medicine; this shrub is very good medicine.]
aambeiebos: Corneels 'Kaldei' Christiaan (Spoegrivier, Namaqualand)	'*Hy groei meer in die nat wêreld en word gebruik vir aambeie*' [It grows in moist places and is used to treat haemorrhoids.]
t`ôrrôbos* or *t`ôrrôboegoe: Esau Flink (Vanrhynsdorp)	'*Trek af vir griep die blare en die stokkies*' [Use the leaves and the twigs as an infusion for the treatment of influenza.]

'–' indicates no information

Table 3: Compounds of six samples of *Pteronia camphorata* as identified by GC–MS (percentage area)

	Locality	Wupperthal			Montagu Pass		
	Samples	**1**	**2**	**3**	**1**	**2**	**3**
	Yield (% dry weight)	**0.53**	**0.59**	**0.71**	**NA**	**NA**	**NA**
RRI	**Major compounds**						
1000	Decane	0.3	0.5	1.9	–	–	–
1032	α-Pinene	0.3	0.6	0.5	0.5	0.2	0.7
1118	β-Pinene	2.1	3.1	2.0	1.1	0.7	0.9
1132	**Sabinene**	**3.5**	**3.6**	**5.1**	**9.1**	**7.1**	**12.7**
1174	Myrcene	6.0	1.6	2.5	–	–	–
1176	α-Phellandrene	–	–	–	1.6	1.7	5.5
1188	α-Terpinene	0.3	0.2	0.3	–	0.1	0.3
1195	Dihydro1,8-cineole	–	–	–	–	0.1	0.1
1203	**Limonene**	**9.9**	**17.1**	**2.0**	**5.0**	**3.8**	**7.7**
1213	**1,8-Cineole**	**22.3**	**17.2**	**26.5**	**42.7**	**40.4**	**42.6**
1246	(*Z*)-β-Ocimene	4.6	5.5	4.6	1.0	0.8	1.5
1255	γ-Terpinene	0.8	0.4	0.6	1.2	0.4	0.7
1266	(*E*)-β-Ocimene	2.6	2.7	2.3	0.2	0.03	0.1
1280	***p*-Cymene**	**3.5**	**15.7**	**25.6**	**17.1**	**21.1**	**10.0**
1290	Terpinolene	–	–	–	0.3	–	0.1
1400	Tetradecane	0.9	5.4	4.1	–	–	–
1450	*trans*-Linalool oxide (furanoid)	–	–	–	–	–	0.1
1474	*trans*-Sabinene hydrate	–	–	–	0.4	0.8	0.6
1450+	*cis*-Linalool oxide (furanoid)	–	–	–	–	0.1	0.1
1500	Pentadecane	–	5.7	4.2	–	–	–
1512	Dillether	–	–	–	–	0.04	–
1553	Linalool	–	–	–	0.4	2.8	3.1
1571	*trans*-*p*-Menth-2-en-1-ol	–	–	–	0.3	0.1	0.2
1586	Pinocarvone	–	–	–	0.1	–	–
1600	Hexadecane	1.0	5.4	4.2	–	–	–
1611	**Terpinen-4-ol**	**5.3**	**2.3**	**3.6**	**5.8**	**3.1**	**2.4**
1629	*cis*-α-Bisabolene	0.2	0.3	–	–	–	–
1632	*cis*-*p*-Menth-2-en-1-ol	–	–	–	–	–	0.2
1639	*trans*-*p*-Menth-2,8-dien-1-ol	–	–	–	0.4	0.4	–
1648	Myrtenal	–	–	–	0.1	0.2	0.1
1651	Sabinaketone	–	–	–	0.3	0.2	0.2
1661	*trans*-Pinocarveol	–	–	–	0.2	0.2	0.1
1662	Estragol	1.0	0.7	0.4	–	–	–
1671	**Methyl chavicol (estragol)**	–	–	–	**4.5**	**6.8**	**2.9**
1682	α-Terpineol	–	–	–	2.3	4.8	3.9
1700	Limonene-4-ol	–	–	–	0.05	0.4	–
1700	Heptadecane	1.0	4.3	3.4	–	–	–
1751	Carvone	0.2	0.7	–	1.2	1	0.7
1755	Bicyclogermacrene	1.0	0.6	0.4	–	–	–
1765	Geranyl acetone	–	–	–	–	0.1	0.1
1773	δ-Cadinene	–	–	–	0.1	–	0.1
1776	γ-Cadinene	–	–	–	–	0.1	–

Table 3 continued

	Locality	Wupperthal			Montagu Pass		
	Samples	**1**	**2**	**3**	**1**	**2**	**3**
1797	*p*-Methyl-acetophenone	–	–	–	0.5	0.5	–
1800	Octadecane	0.3	1.4	1.2	–	–	–
1802	Cumin aldehyde	–	–	–	0.4	0.4	0.2
1804	Myrtenol	–	–	–	–	0.2	0.1
1814	*p*-Mentha-1,5-dien-7-ol	–	–	–	0.2	–	–
1834	*trans*-Carveol	–	–	–	0.4	0.3	0.2
1854	Germacrene B	–	0.4	–	–	–	–
1864	*p*-Cymen-8-ol	–	0.1	0.4	0.2	0.4	0.2
1882	*cis*-Carveol	–	–	–	0.1	–	0.1
2008	Caryophyllene oxide	–	–	–	0.3	0.3	0.4
2012	1-Allyl-2,4-dimethoxybenzene	–	–	–	0.3	–	–
2030	**Methyl eugenol**	**23.4**	0.5	–	–	0.4	0.2
2202	Germacrene D-4-ol	–	–	–	–	–	0.1
2081	Humulene-epoxide III	–	–	–	–	0.04	–
2113	Cumin alcohol	–	–	–	0.2	0.1	0.1
2144	Spathulenol	–	–	–	0.1	–	0.1
2187	T-Cadinol	0.2	–	–	–	–	0.04
2228	Himachalol	–	–	–	0.1	0.1	0.2
2255	α-Cadinol	–	–	–	0.2	–	–
Total		**90.7**	**96.0**	**95.8**	**99.4**	**99.3**	**99.7**

Note: *Three essential oil samples of P. camphorata from Kleinvlei, Wupperthal (P. camphorata var. armata) are compared with published results for three samples of P. camphorata var. camphorata from Montagu Pass.[16] The three specimens from each population are shown as 1, 2 and 3.*

Table 4: Minimum inhibitory concentrations for extracts and essential oils of three specimens of *Pteronia camphorata* var. *armata*

Extract	**Sample number**	**Minimum inhibitory concentration (mg/ml)**			
		***Moraxella catarrhalis* ATCC 23246**	***Mycobac-terium smegmatis* ATCC 14468**	***Crypto-coccus neoformans* ATCC 90112**	***Klebsiella pneumoniae* ATCC 13883**
H_2O extract	1	> 8	3.0	**0.5**	8.0
H_2O extract	2	> 8	2.0	> 8	8.0
H_2O extract	3	> 8	1.5	> 8	8.0
MeOH:H_2O extract	1	> 8	8.0	**0.2**	> 8
MeOH:H_2O extract	2	> 8	8.0	**0.1**	> 8
MeOH:H_2O extract	3	> 8	1.5	**0.1**	> 8
MeOH:CH_2Cl_2 extract	1	4.0	**0.5**	**0.8**	4.0
MeOH:CH_2Cl_2 extract	2	> 8	**0.3**	**0.8**	4.0
MeOH:CH_2Cl_2 extract	3	> 8	**0.5**	**0.8**	1.5
Essential oil	1	> 16	**1.0**	**0.3**	4.0
Essential oil	2	> 16	**1.0**	**0.3**	4.0
Essential oil	3	> 16	**1.0**	**0.5**	4.0
Positive control (ciprofloxacin/amphotericin B)		0.313 μg	0.313 μg	2.5 μg	0.078 μg

Note: Concentrations were tested on a selection of pathogens associated with respiratory infections, including a yeast (*Cryptococcus neoformans*), two Gram-negative bacteria (*Moraxella catarrhalis* and *Klebsiella pneumoniae*) and one Gram-positive bacterium (*Mycobacterium smegmatis*).

This word appears to be derived from a corruption of the word '*|nhora*' (| denoting the dental click) and *boegoe* (the common Khoisan term for aromatic bushes). It was explained to B.E.V.W. that the term *|nhora* in the San culture refers to the 'life force' in human beings, which is believed to be situated in the nape of the neck. Directly translated the name means 'life force buchu', reflecting the perceived importance of this plant species.

Recent enquiries in various parts of the Western and Northern Cape provinces (including the Cederberg region, where the species is quite common) revealed only a few known contemporary uses. In the Kamiesberg area of Namaqualand in Northern Cape, we located two herbalists with original knowledge about the species and its uses in the Nama culture. The late Anna 'Boom' Stewe was a herbalist from Leliefontein in the Kamiesberg, who obtained her medicinal knowledge from the late Jan 'Bordhoed' Beukes, a local *bossiedokter* – a term of honour for an acknowledged herb doctor. The plant, which Anna Stewe called *|nôrraboegoe*, is used in a mixture with *org* [*Notobubon pearsonii* (Adamson) Magee] and *jantjie-bêrend* [*Sutherlandia frutescens* (L.) R. Br. which is] [=*Lessertia frutescens* (L.) Goldblatt & J.C. Manning] to treat tuberculosis. A small glass (*kelkie*) of a hot water infusion (tea) of the leaves of the three species is taken three times a day, and is considered to be the best medicine for tuberculosis. This medicine was used by an elderly lady, whose name is no longer recalled, for convalescent tuberculosis patients. Powdered leaves mixed with other herbs can also be used medicinally as snuff, and infusions of a leafy twig in hot milk or water are used to alleviate toothache. Powdered leaves are applied to a piece of cotton wool and inserted into the ear for relief of earache (*dit het 'n pyndodende effek* – 'it has a painkilling effect'). The mixture can also be used for flatulence ('to remove wind'). The use against earache and flatulence was confirmed by three other local inhabitants of Kamiesberg, and another person added that an infusion can be used to treat rheumatism.[8]

The second local expert in the Kamiesberg was the late Gert 'Joelk' Dirkse, a herbal doctor from Paulshoek and one of the last professional *bossiedokters* in Namaqualand. He independently supplied information about *P. camphorata*. Powdered root is used to treat febrile convulsions in infants, and powdered leaves are used to treat convulsions and epilepsy in adults as well as for blood purification. Gert Dirkse stated that a decoction of the leaves, mixed with *baarbos* (*Limeum africanum* L.) and *slangbos* [*Stoebe plumosa* (L.) Thunb.] is taken orally and is very effective in treating general malaise (*siek voel*).

We also heard recent anecdotes from the small Namaqualand village of Spoegrivier. An experienced and highly knowledgeable midwife, Ouma Sarah 'Toesie' Fortuin, still remembered *|ôrrôboegoe* or *|ôrrôbosboegoe*. According to Sarah Fortuin, *|ôrrô* refers to the back of the neck, and the plant is used to treat inflammation and redness of the neck in children. It is also a medicine for chest ailments, and is considered highly effective (*hierdie bos is baie goeie medisyne* – 'this shrub is very good medicine'). Corneels 'Kaldei' Christiaan from the same village reported that the plant grows only in moist areas and is used for the treatment of haemorrhoids – hence his vernacular name for the plant, *aambeiebos* (haemorrhoids bush). Further south, near Vanrhynsdorp not far from Nieuwoudtville, Esau Flink reported that he uses a decoction of the leaves and twigs to treat influenza. He referred to it as *t`ôrrôbos.*

Given its obvious medicinal importance, it is surprising that the uses of *P. camphorata* have remained almost unknown to science until recently. We were fortunate to have the opportunity to record, and preserve for posterity, the rich knowledge of two very knowledgeable and experienced herbalists, Anna Stewe and Gert Dirkse. Sadly, neither of them had any apprentices to whom they could pass on their knowledge.

Anatomy

The leaves of *P. camphorata* are amphistomatic. They have a thin cuticle and the outer periclinal cell walls of the epidermal cells are markedly thickened. The mesophyll is composed of usually two layers of palisade parenchyma, more highly developed adaxially than abaxially, which surround the central spongy parenchyma (Figure 2). Secretory structures that can be referred to as oil glands or cavities occur in the spongy parenchyma adjacent to the phloem of some vascular bundles. Macroscopically they are visible as translucent dots (Figure 1). Secretory trichomes are situated in the medial adaxial groove, but can sometimes also be found in the corresponding position on the abaxial surface.

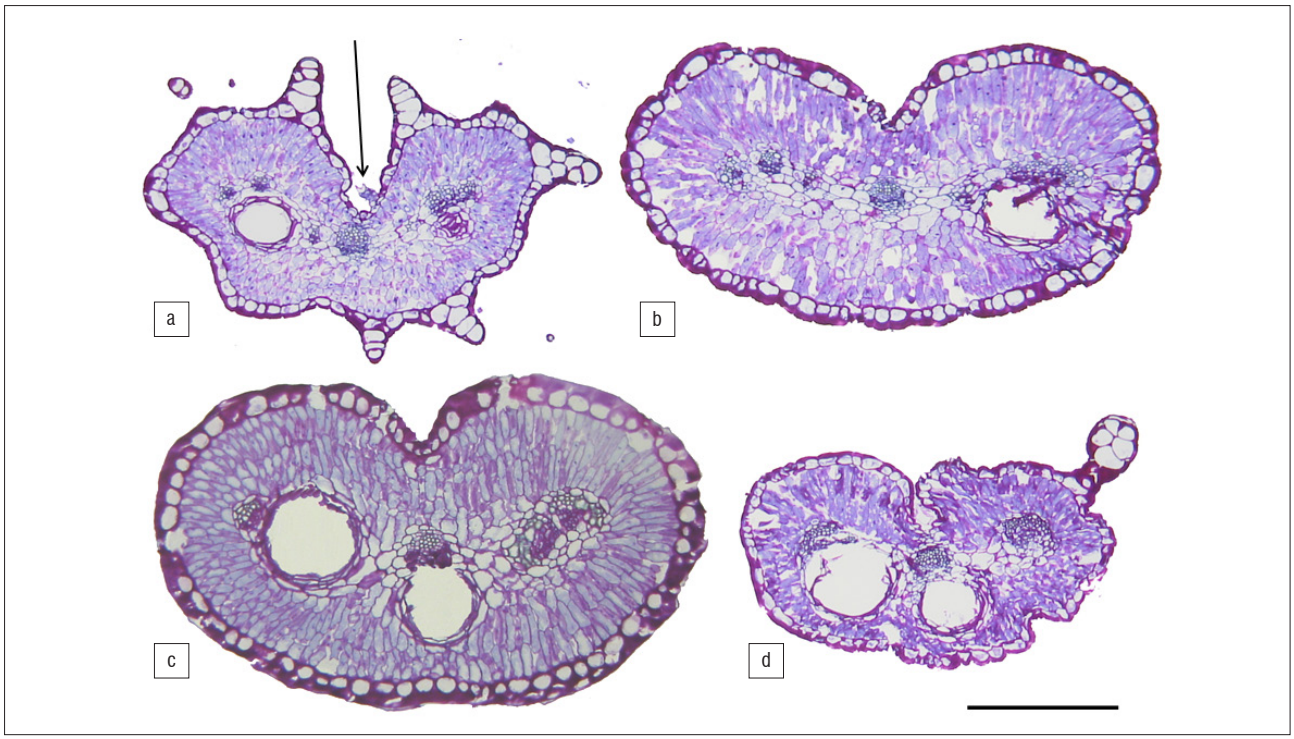


Figure 2: Transverse sections of leaves from four varieties of *Pteronia camphorata*. (a) *P. camphorata* var. *armata*, (b) *P. camphorata* var. *camphorata*, (c) *P. camphorata* var. *laevigata*, and (d) *P. camphorata* var. *longifolia*. Oil glands are present in all, and secretory trichomes in (a) and (d).

Key: The arrow indicates a secretory trichome; scale bar = 0.4 mm.

It appears that three types of secretory structures are present in the genus *Pteronia*, and the ethnobotanically relevant species can be classified into four groups based on their combinations. The structures are as follows: (1) globose oil glands or cavities, similar to those in *Citrus,* which are present in *P. onobromoides* D.C.*, P. stricta* Aiton and *P. succulent* Thunb.; (2) oil ducts associated with the vascular bundles, mainly the midrib, which are present in *P. adenocarpa* Harv.; (3) oil ducts associated with the vascular bundles as well as secretory trichomes, which occur in *P. cinerea* L.f., *P. divaricata* Less., *P. incana* D.C. and *P. lucillioides* D.C.; and (4) globose oil glands and secretory trichomes mainly in the medial grooves – this type is present in *P. camphorata*. The leaf anatomy of *Pteronia* species appears to be of considerable value in providing diagnostic characters, and has the potential to contribute to a better understanding of the infrageneric taxonomy and relationships of the genus.

Essential oil composition

We identified 21 to 25 volatile components in the three samples of *P. camphorata* var. *armata* that we analysed from Wupperthal (Table 3). A similar combination of main compounds was found in *P. camphorata* var. *camphorata* from Montagu Pass in the southern Cape, as studied by Coovadia[16] and reported by Viljoen et al.[17] The major compounds in one or both of the varieties are several monoterpenes as well as sesquiterpenes (Table 3). The two varieties agree in the presence of sabinene, limonene, 1,8-cineole, *p*-cymene and terpinene-4-ol as main compounds. In addition, (*Z*)-β-ocimene is a major constituent in Wupperthal samples but a minor constituent in Montagu Pass samples. Estragol occurs at relatively high levels at Montagu but only as minor compounds at Wupperthal. Methyl eugenol was a major compound in one of the Wupperthal samples. Both varieties have smaller amounts of α-pinene, β-pinene, γ-terpinene and (*E*)-β-ocimene.

A number of non-volatile diterpenes and other phenolic compounds have been isolated from *P. camphorata* and other *Pteronia* species.[18] These include 3,4-dimethoxypropiophenone, eugenol methyl ether and several other more widespread compounds. The biological, chemotaxonomic and medicinal significance of these compounds are as yet unknown.

We compared the main essential oil compounds in *P. camphorata* with those of four other ethnobotanically relevant species for which data are available.[16] The results showed the following general patterns: (1) sabinene, limonene, 1,8-cineole, *p*-cymene and myrcene are main constituents in most of the species, and the same pattern is evident for *P. camphorata*; (2) β-pinene is a major compound in *P. adenocarpa* and *P. incana* and is present as a minor compound in all other species; and (3) sesquiterpenoids appear to be more restricted in their distribution. There might well be a link between the essential oil composition and antimicrobial activity. A previous study showed that cineole and limonene may interact synergistically to exert enhanced antimicrobial activity.[19]

Antimicrobial activity

The results of our antimicrobial study on the extracts of *P. camphorata* are shown in Table 4, with noteworthy activity highlighted in bold text. Noteworthy activities for extracts were considered where MIC values were below 1.0 mg/ml[20–21] and for essential oil samples where MIC values were below 2.0 mg/ml[22]. The methanol-to-dichloromethane (MeOH : CH_2Cl_2) extracts were the most active against *Mycobacterium smegmatis*, with MIC values as low as 0.3 mg/ml to 0.5 mg/ml. The methanol-to-water (MeOH : H_2O) extracts were the most active against *C. neoformans*, with MIC values as low as 0.1 mg/ml to 0.2 mg/ml. The essential oils also showed strongest antimicrobial activity against *C. neoformans,* with the most noteworthy having an MIC value of 0.3 mg/ml.

Water and methanol-to-water extracts generally showed poor to no activity at the highest concentration tested, against all organisms studied except for *C. neoformans*. In the latter case, the mean MIC value was 0.5 (H_2O extract), with MIC values ranging from 1.5 mg/ml to 3.0 mg/ml (H_2O extracts) for the various *Mycobacterium* species tested.

Conclusion

P. camphorata is poorly known as a medicinal plant, but available information and recently recorded ethnobotanical data suggest it was once an important Khoisan remedy. Almost no local knowledge or local users could be found in most of the rural localities where the plants grow. The traditional uses include treatment of respiratory conditions (colds, fever, influenza, chest ailments and tuberculosis) and inflammation of the neck, convulsions and haemorrhoids. These uses suggest the plant has antimicrobial, anti-inflammatory and antipyretic properties.

The globose oil glands found in *P. camphorata* are similar to those of *P. onobromoides*, *P. stricta* and *P. succulenta* but not those of any other species. *P. camphorata* differs from the others by the presence of additional glandular trichomes in the medial grooves. The leaf anatomy appears to have some diagnostic value at the species level, and holds considerable potential as a source of taxonomic characters.

P. camphorata essential oil has a similar combination of main compounds as that found in other *Pteronia* species: sabinene, limonene, 1,8-cineole, *p*-cymene and terpinene-4-ol are the main constituents. Smaller amounts of α-pinene, β-pinene, (*Z*)-β-ocimene, γ-terpinene, (*E*)-β-ocimene and myrcene also occur in all or most of the samples of this species. The plant has small amounts of sesquiterpenes, such as bicyclogermacrene – which is a main compound in *P. divaricata* and a minor compound in *P. incana*.

Most of the extracts, as well as the essential oil of the plant, display activity against *C. neoformans*. The dichloromethane extracts and essential oil samples were active against *M. smegmatis*. The level of activity against these respiratory pathogens seems to support the reported efficacy of the traditional treatment against tuberculosis, colds and influenza.

Authors' contributions

I.M.H. prepared the samples and performed the experiments; P.M.T. provided anatomical know-how and infrastructure, and prepared the anatomical descriptions; S.F.V.V. was responsible for the experimental design of antimicrobial tests, and provided the microbial samples and infrastructure; G.P.P.K. performed the GC–MS analyses and calculations; J.M.N. contributed original ethnobotanical data; A.M.V. made conceptual contributions, guided the essential oil analyses and provided the analytical infrastructure; and B.E.V.W. was the project leader, conceptualised the study, wrote the first draft and contributed taxonomic and ethnobotanical data.

Acknowledgements

Financial support from the National Research Foundation is gratefully acknowledged. We thank and honour Mr Willem 'Blikkies' Steenkamp of the farm Oorlogskloof near Nieuwoudtville; the late Ms Anna 'Boom' Stewe of Leliefontein; the late Mr Gert 'Joelk' Dirkse of Paulshoek; Mrs Sarah 'Toesie' Fortuin and Corneels 'Kaldei' Christiaan of Spoegrivier; and Mr Esau Flink of Vanrhynsdorp, for sharing their remarkable traditional knowledge on medicinal plants. We also thank Mr Paul Herman at the South African National Biodiversity Institute in Pretoria for kindly identifying the four varieties of *P. camphorata* used in this study.

References

1. Laidler PW. The magic medicine of the Hottentots. S Afr J Sci. 1928;25:433–447.
2. Nortje JM, Van Wyk B-E. Medicinal ethnobotany of the Kamiesberg, Namaqualand, South Africa. J Ethnopharmacol. 2015;171:205–222. http://dx.doi.org/10.1016/j.jep.2015.04.049
3. Hutchinson J, Phillips EP. A revision of the genus *Pteronia* (Compositae). Ann S Afr Mus. 1917;9:277–329.
4. Hulley IM, Viljoen AM, Tilney PM, Van Vuuren SF, Kamatou GPP, Van Wyk B-E. Ethnobotany, leaf anatomy, essential oil composition and antibacterial activity of *Pteronia onobromoides* (Asteraceae). S Afr J Bot. 2010;76:43–48. http://dx.doi.org/10.1016/j.sajb.2009.06.012

5. Hulley IM, Viljoen AM, Tilney PM, Van Vuuren SF, Kamatou GPP, Van Wyk B-E. The ethnobotany, leaf anatomy, essential oil variation and biological activity of *Pteronia incana* (Asteraceae). S Afr J Bot. 2010;76:668–675. http://dx.doi.org/10.1016/j.sajb.2010.08.007
6. Hulley IM, Viljoen AM, Tilney PM, Van Vuuren SF, Kamatou GPP, Van Wyk B-E. *Pteronia divaricata* (Asteraceae): A newly recorded Cape herbal medicine. S Afr J Bot. 2011;77:66–74. http://dx.doi.org/10.1016/j.sajb.2010.05.010
7. International Society of Ethnobiology. International Society of Ethnobiology Code of Ethics. Bristol, Vermont; 2006 (with 2008 additions). [Cited 2015 May 6]. Available from: http://ethnobiology.net/code-of-ethics/
8. Nortje JM. Medicinal ethnobotany of the Kamiesberg, Namaqualand, Northern Cape Province, South Africa [MSc dissertation]. Johannesburg: University of Johannesburg; 2011.
9. Marloth R. The Flora of South Africa. Volume 3. Cape Town: Darter Bros; 1932. p. 250.
10. Smith CA. Common names of South African plants. Memoirs of the Botanical Survey of South Africa 35. Pretoria: Department of Agricultural Technical Services; 1966.
11. Burman L, Bean A. Hottentots Holland to Hermanus. Botanical Society of South Africa: Kirstenbosch; 1985.
12. Watt JM, Breyer-Brandwijk MG. The medicinal and poisonous plants of southern and eastern Africa. 2nd ed. London: Livingstone; 1962. p. 255.
13. Kling H. Die sieketrooster: 'n Noodanker in tyd van siekte en ongevalle. Cape Town: Van de Sandt de Villiers; 1923.
14. Arnold TH, Prentice CA, Hawker LC, Snyman EE, Tomalin M, Crouch NR, et al. Medicinal and magical plants of southern Africa: An annotated checklist. Strelitzia, Volume 13. Pretoria: National Botanical Institute; 2002. p. 37.
15. Van Wyk B-E, Gericke N. People's plants: A guide to useful plants of Southern Africa. Pretoria: Briza Publications; 2000. p. 218, p. 224.
16. Coovadia ZH. The antimicrobial properties and chemical composition of leaf extracts and essential oils of indigenous *Pteronia* species [MSc dissertation]. Johannesburg: University of the Witwatersrand; 2007.
17. Viljoen AM, Kamatou GPP, Coovadia ZH, Özek T, Başer KHC. Rare sesquiterpenes from South African *Pteronia* species. S Afr J Bot. 2010;76:146–152. http://dx.doi.org/10.1016/j.sajb.2009.07.012
18. Zdero C, Jakupovic J, Bohlmann F. Diterpenes and other constituents from *Pteronia* species. Phytochemistry. 1990;29:1231–1245. http://dx.doi.org/10.1016/0031-9422(90)85434-H
19. Van Vuuren SF, Viljoen AM. Antimicrobial activity of limonene enantiomers and 1,8-cineole alone and in combination. Flavour Fragr J. 2007;22:540–544. http://dx.doi.org/10.1002/ffj.1843
20. Gibbons S. Anti-staphylococcal plant natural products. Nat Prod Rep. 2004;21:263–277. http://dx.doi.org/10.1039/b212695h
21. Rios JL, Recio MC. Medicinal plants and antimicrobial activity. J Ethnopharmacol. 2005;100:80–84. http://dx.doi.org/10.1016/j.jep.2005.04.025
22. Van Vuuren SF. Antimicrobial activity of South African medicinal plants. J Ethnopharmacol. 2008;119:462–472. http://dx.doi.org/10.1016/j.jep.2008.05.038

Implications of summer breeding frogs from Langebaanweg, South Africa: Regional climate evolution at 5.1 mya

AUTHORS:
Thalassa Matthews[1]
G. John Measey[2]
†David L. Roberts[3]
†*Deceased*

AFFILIATIONS:
[1]DST–NRF Centre of Excellence for Palaeontology, Iziko South African Museum, Cape Town, South Africa
[2]Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa
[3]Department of Geography, University of the Free State, Bloemfontein, South Africa

CORRESPONDENCE TO:
Thalassa Matthews

EMAIL:
tmatthews@iziko.org.za

DATES:
Received: 29 Feb. 2016
Revised: 12 May 2016
Accepted: 16 May 2016

KEYWORDS:
palaeoclimate; *Ptychadena*; Summer Rainfall Zone; Early Pliocene; fossil frogs

HOW TO CITE:
Matthews T, Measey GJ, Roberts DL. Implications of summer breeding frogs from Langebaanweg, South Africa: Regional climate evolution at 5.1 mya. S Afr J Sci. 2016;112(9/10), Art. #2016-0070, 7 pages. http://dx.doi.org/10.17159/sajs.2016/20160070

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
DST–NRF Centre of Excellence in Palaeosciences; DST–NRF Centre of Excellence for Invasion Biology; National Research Foundation (South Africa); University of the Free State

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

No direct palaeoclimatic proxies have been available to indicate the seasonality or amount of rainfall on the west coast of southern Africa during the Early Pliocene. The Benguela Upwelling System (BUS) is today one of the factors responsible for the present-day aridity on the west coast of southern Africa. The initiation of the BUS is frequently linked to the entrenchment of aridity and the establishment of the current winter rainfall pattern on the west coast; however, marine proxies are inconclusive regarding the effects of past fluctuations in the BUS and sea surface temperatures on the rainfall regime. Neither the fossil evidence nor the fact that plants using the C_3 photosynthetic pathway predominate at this time, provide direct evidence of winter rainfall at Langebaanweg. We challenge certain assumptions which are commonly made in the literature regarding the timing of inception of a winter rainfall regime on the west coast and the onset of aridity in the Langebaan region, and provide new evidence as to seasonality of rainfall at Langebaanweg in the Early Pliocene. Herein, the identification of frog species from the genus *Ptychadena* from Langebaanweg provides new and compelling evidence for a summer rainfall regime, or of at least significant summer rainfall, at 5.1 mya in the southwestern Cape of South Africa.

Significance:

- Advances understanding of the evolution of the winter rainfall zone on the west coast of South Africa
- Assesses evidence for the inception of aridity and a winter rainfall regime on the west coast of South Africa
- Fossil Ptychadenidae from the Early Pliocene site of Langebaanweg provide evidence for a summer rainfall regime at 5.1 mya on the west coast of the southwestern Cape.

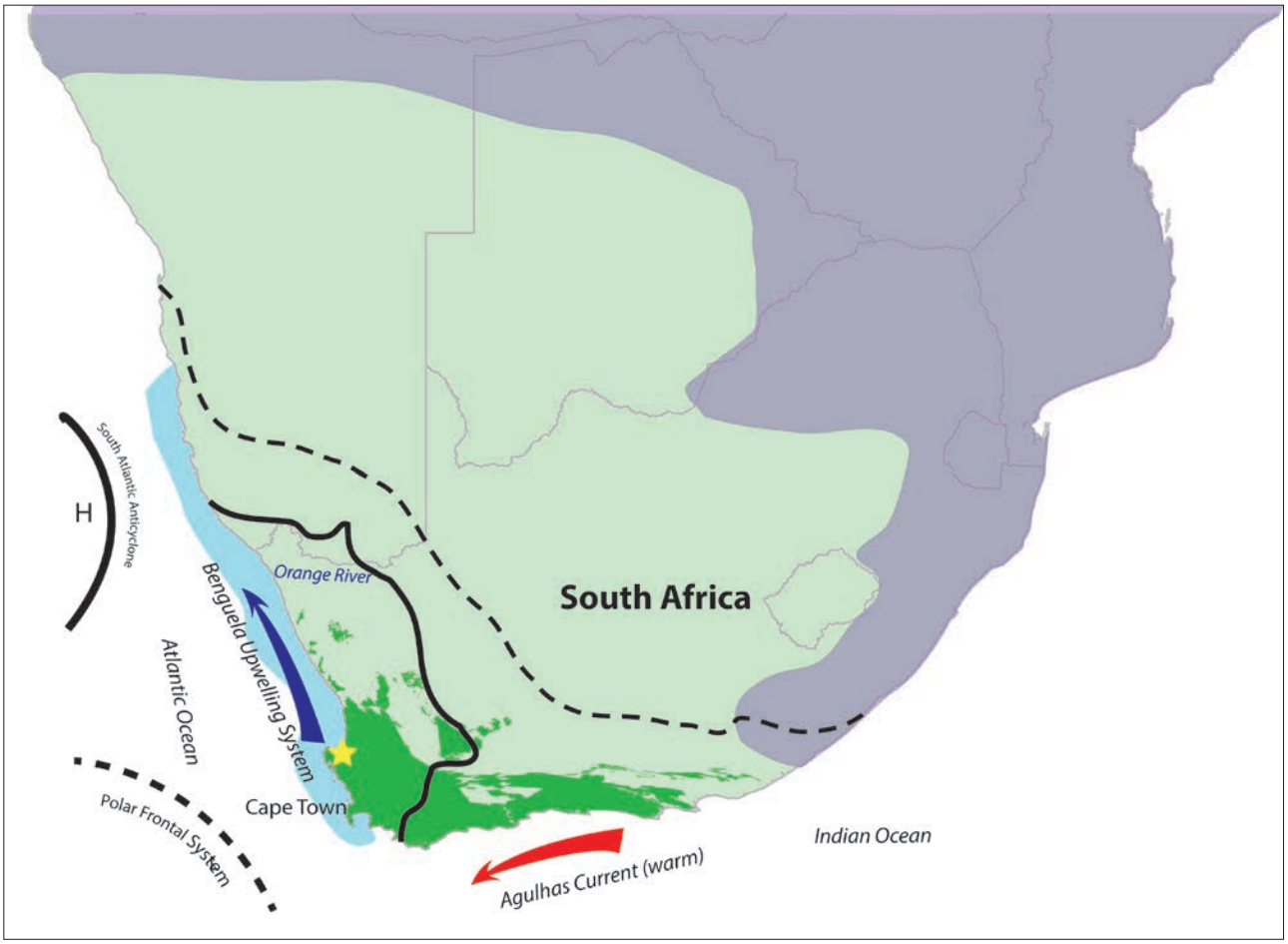
Introduction

Southern Africa is a predominantly summer rainfall region but the Winter Rainfall Zone (WRZ) situated along the southwestern and southern tip of the African continent, lying at the boundary between southern hemisphere temperate and tropical climate systems, is an exception (Figure 1). Evidence for the inception, geographical extent, intensity and fluctuations of the WRZ prior to the last glacial period is currently fragmented through both time and space, and a coherent understanding of the evolution of this ecosystem, unique in sub-Saharan Africa, has not been attained.[1] Although progress has been made in understanding climate change and rainfall seasonality in the Late Quaternary,[2-5] the fluctuations of these variables over the Neogene are complex and incompletely understood. The Neogene terrestrial fossil record is sporadic and incomplete as a result of the lack of preservation of terrestrial organic materials and thus direct climatic proxies, deposits which are problematic in terms of dating, and a lack of understanding of the geomorphological evolution of the subcontinent.[5,6] Much therefore remains to be elucidated about the inception and evolution of the WRZ. In this paper, we challenge certain assumptions which are commonly made in the literature regarding the inception of a winter rainfall regime on the west coast and the onset of aridity in the Langebaan region, and provide new evidence for rainfall seasonality at Langebaanweg in the Early Pliocene.

The Early Pliocene (~5.1 million years ago (mya)) site of Langebaanweg (LBW) situated on the southwestern coast of South Africa (Figure 1), is a remarkably rich and world-renowned vertebrate fossil site, representing a time period when fossil assemblages are particularly rare in sub-Saharan Africa.[7] LBW is the site of first appearance in the fossil record for numerous taxa, including large and small mammals, birds, reptiles and amphibians.[7-10] LBW sheds light on the critical period during which tectonism led to the latitudinal migration of climate belts, which resulted in widespread reorganisation in atmospheric and ocean circulation.[11] These events were characterised by a more humid episode, with modern aridity becoming established only in the later Pliocene, between ~4 mya and 2.8 mya.[11] Congruently, the fauna at LBW indicates more humid conditions than at present, but the precise nature of the vegetation in the area remains uncertain and until recent research on the frog fauna there were no effective proxies indicating the amount of rainfall, or seasonality, at the fossil site.[10,12-14]

A recent study of the rich and diverse anuran (frog) community from LBW served as an effective and direct climatic proxy, and showed that this group, sensitive as they are to rainfall and moisture, supported the higher humidity inferred from studies of other faunal groups from the site, and indicated relatively high rainfall.[12,15] The huge contrast between the frog community in the Langebaanweg region today, as compared with 5.1 mya, is clearly indicated by the fact that species richness on the relatively dry west coast of South Africa is low today (1–10 species)[16,17], whereas at LBW, six families and some 19 taxa have been differentiated[12]. The LBW frog fauna is put in context when compared to modern species richness elsewhere in the southwestern Cape, which varies from 11 to 30 species.[16,18]

Currently, southern African Anura are distributed throughout three different rainfall zones: the Summer Rainfall Zone (SRZ), the Winter Rainfall Zone (WRZ) and the Year-round Rainfall Zone (YRZ) which is found on the South African south coast. A number of amphibian taxa from the WRZ and the SRZ overlap in the YRZ.[18]



Note: Grey areas indicate the current distribution of Ptychadena *taxa; the star indicates the position of Langebaanweg. The area to the left of the solid line indicates the current Winter Rainfall Zone and the area to the right between the solid and stipled lines indicates the current Year-round Rainfall Zone. To the right of the stipled line is the Summer Rainfall Zone. The Fynbos Biome is indicated in green. Rainfall zones after Chase and Meadows[3].*

Figure 1: Locality of the fossil site of Langebaanweg and current climatic and oceanographic conditions at the southern tip of Africa.

All frog families and genera previously identified at LBW were ambiguous in terms of the seasonality of rainfall as all but one of the species (a SRZ taxon) identified currently occur within all three rainfall regimes (Table 1). In the present study, we take a step further, utilising the recent identification at LBW of two species of the genus *Ptychadena* (family Ptychadenidae) to shed further light on the seasonality of rainfall at this site. Additionally, this discovery further elucidates the biogeography of the family Ptychadenidae.

Climatic history of the Winter Rainfall Zone

The WRZ receives predominantly winter rainfall from eastward migrating cold fronts embedded in polar cyclonic systems originating over the South Atlantic.[19,20] During summer, the South Atlantic Anticyclone is well developed and migrates to the south, blocking both the westward propagation of easterly waves that bring summer rainfall to much of southern Africa, and the polar frontal systems. The influence of the polar frontal systems diminishes northwards and is linked to decreasing rainfall and increasing aridity.[14] The cold, nutrient-rich Benguela Upwelling System (BUS) supports a diverse and rich range of marine life along the southwestern African coast.[21] The South Atlantic Anticyclone promotes the BUS and is thus the main cause of the present aridity along the southwestern African coast, the Namib, and the dryness of the adjacent interior of the sub-continent.[5,20,22-25] In general, the climate on the west coast is determined by the complex seasonal shifts and interplay of the prevailing winds over the Benguela region.[5] Fluctuations in summer rainfall are linked to upwelling intensity in the BUS[26] and along the south coast[5]. The situation is complex, as indicated by the fact that Chase et al.[5] found that at a hyrax-midden site falling within the present WRZ, but close to the border with the SRZ, summer rainfall variability over the Holocene during the last 7000 years had the greatest impact on water availability, and was the primary determinant of humidity. Analysis of herbivore tooth enamel indicates that there were incursions of C_4 vegetation in the mid-Pleistocene,[27] which could be related to decreased atmospheric pCO_2 conditions during interglacial periods, or possibly to changes in rainfall seasonality.

Recent research on marine proxies for sea surface temperature (SST) in the southern Cape Basin – such as calcareous dinoflagellate cysts and alkenones – suggests the inception of the BUS at about 10.5–10 mya.[28,29] How the BUS became established is comprehensively summarised in Neumann and Bamford[6]. Some authors[29,30] have suggested that tectonic uplift in southwestern Africa would have led to a cooling in the BUS at 12 mya, and after 5 mya, but the timing and cause of such tectonic events, and even their occurrence, is controversial[6,31]. The initiation of the BUS has frequently been linked in the literature to summer aridity and the entrenchment of the current winter rainfall pattern on the west coast[6,24,32,33], and this in turn has been taken to have influenced other processes, such as the diversification of plants (see Altwegg et al.[33] – although a recent analysis[1] failed to find proof that seasonal aridity at ~8 mya triggered floristic radiation and diversification in the Cape Floral Region).

The connection between the BUS and seasonality of rainfall is unclear as estimates of the evolution of SSTs are not available[34] and the marine record indicates that the BUS and SSTs have fluctuated considerably over time in the late Neogene. High glacio-eustatic sea levels and the presence of warm-water molluscan taxa in the middle to late Pliocene[35,36] add further complexity to deciphering the evolution of oceanic and climatic conditions in the current WRZ.

Table 1: Distribution of extant frog families and genera according to the current rainfall regime. Fossil taxa found at Langebaanweg are bolded.

Family	Genera		
	Summer rainfall	Winter rainfall	Year-round rainfall
Arthroleptidae			
	Arthroleptis	–	–
	Leptopelis	–	–
Hyperoliidae			
	Hyperolius	***Hyperolius***	***Hyperolius***
	Afrixalus	–	*Afrixalus*
	Semnodactylus	*Semnodactylus*	*Semnodactylus*
	Kassina	–	–
Heleophrynidae			
	Heleophryne	***Heleophryne***	***Heleophryne***
	Hadromophryne	–	–
Brevicipitidae			
	Breviceps	***Breviceps***	***Breviceps***
Pyxicephalidae			
	Amietia	*Amietia*	*Amietia*
	Anhydrophryne	–	
	Cacosternum	*Cacosternum*	*Cacosternum*
	Strongylopus	*Strongylopus*	*Strongylopus*
	Tomopterna	***Tomopterna***	***Tomopterna***
	Pyxicephalus	–	–
	Natalobatrachus	–	–
	–	*Microbatrachella Arthroleptella*	–
	–	*Poyntonia*	–
Phrynobatrachidae			
	Phrynobatrachus	–	–
Ptychadenidae			
	Ptychadena		
	Hildebrandtia	–	–
	Lanzarana	–	–
Bufonidae			
	Amietophrynus	***Amietophrynus***	***Amietophrynus***
	Vandijkophrynus	*Vandijkophrynus*	*Vandijkophrynus*
	Poyntonophrynus	*Poyntonophrynus*	*Poyntonophrynus*
	Schismaderma	–	–
	–	*Capensibufo*	*Capensibufo*
Pipidae			
	Xenopus	***Xenopus***	***Xenopus***
Microhylidae			
	Phrynomantis	–	–
Hemisotidae		–	–
	Hemisus	–	–
Rhacophoridae		–	–
	Chiromantis	–	–
Families	12	6	6
Genera	29	17	14

There is also evidence of an oceanic subsurface warming between 6.5 mya and 5.0 mya which may be related to variability in the strength of North Atlantic deepwater formation on the temperatures of the intermediate waters of the South Atlantic.[34] The climatic data during LBW times are ambiguous in terms of oceanographic conditions, with SSTs apparently in sharp decline, but with low productivity (suggesting reduced upwelling) indicated by a drop in total organic carbon.[34] The fossil-bearing members of the Varswater Formation present at LBW conversely contain an abundance of authigenic phosphate, which is widely interpreted as being indicative of strong upwelling (Roberts et al.[14] and references therein). The presence of some mollusc taxa from LBW which are currently found northwards from East London and eastwards from False Bay suggest warmer ocean temperatures than present.[37]

Identification and provenance of *Ptychadena* at Langebaanweg

The majority of fossils from LBW were recovered from two highly fossiliferous members which form part of the Varswater Formation, namely the Muishondfontein Pelletal Phosphate Member (MPPM) and the Langeberg Quartzose Sand Member (LQSM). See Roberts et al.[14] for further details on the geology of the site.

An examination of LQSM sediments (~5.1 mya) collected from a mine dump during phosphate-mining operations several decades ago was recently undertaken. The interpretation of the LQSM in the area from which the dump derived (the so-called 'east stream') was that it represented a river floodplain.[38] In the course of this study, an additional anuran genus (accession number SAM-PQL-70839) was identified – *Ptychadena* (Family: Ptychadenidae) – on the basis of a single sacrum (Figure 2). Subsequently a further ptychadenid, which appears to belong to a relatively larger species, was recovered from MPPM deposits which were excavated from the main dig site at the West Coast Fossil Park from square G5 (accession number SAM-PQL-71524) (Figure 2). The larger Ptychadenid showed several morphological differences to the smaller taxon, including a greater distance between the anterior and posterior articular ends of the sacrum (best seen in ventral view, Figure 2b and 2f), a larger neural spine, and a wider spacing of the sacral articular condyles.

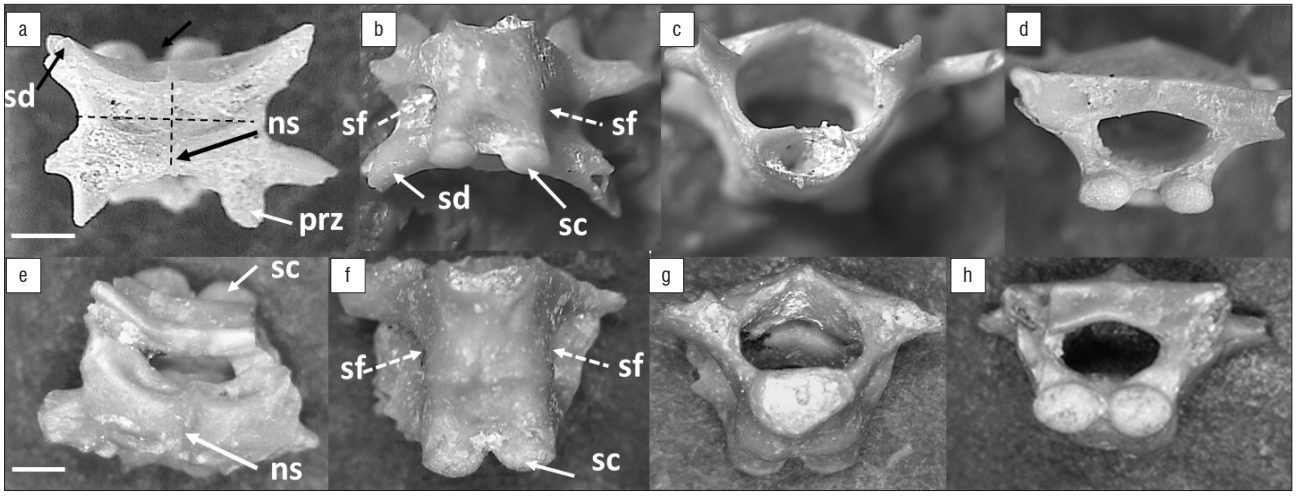
The sacrum of a modern Ptychadenid, *Ptychadena mascareniensis*, from the Iziko South African Museum (ZR-045512), was scanned using microcomputed tomography and then measured for comparative purposes (see Figure 2 for positioning of measurements). This species currently enjoys a wide distribution in western, eastern and central Africa, extending down into South Africa. The breadth of the sacrum was 3.27 mm and the height 1.46 mm. The modern *Ptychadena* thus falls in between the two fossil taxa in terms of size as the smaller of the fossil taxa (SAM-PQL-70839) measured 2.79 mm by 1.39 mm, and the larger (SAM-PQL-71524) 4.60 mm by 1.88 mm for breadth and height, respectively.

Extant Ptychadenidae comprise the genera *Ptychadena, Hildebrandtia* and *Lanzarana*, and are monophyletic.[39] Four unique morphological synapomorphies are noted, including the symmetrical fusion of the eighth presacral vertebra with the sacral vertebra (sacrum), forming a composite sacrum.[40] The LBW ptychadenid sacrum displays the fusion of the presacral vertebra and sacrum, together with the distinctive anterolateral orientation and morphology of the transverse processes of members of the genus *Ptychadena*, including a characteristic foramen above the centrum that serves as the outlet for the spinal nerve (as noted by Blackburn et al.[41]). In addition, *Ptychadena* species differ from many ranids in that the neural arches of the composite sacrum lack deep grooves and fossae on the dorsal surface and exhibit a single minute low-lying neural spine.[40] Blackburn et al.[41] provide conclusive evidence, through a thorough investigation of extant and fossil taxa, that the Ptychadenidae are the only major clade of frogs with the above sacral morphologies, making the identification of the LBW taxa indisputable.

LBW extends our knowledge of the fossil distribution of ptychadenids to include the southwestern Cape at 5.1 mya, and represents the most southerly occurrence of this genus to date in the African fossil record. The fossil frog bones reported on in this paper are curated by the Iziko South African Museum, Cape Town, and each bone reported on has been allocated an identifying accession number.

Other fossil sites containing Ptychadena *species*

The antiquity of the *Ptychadena* lineage is indicated by the earliest appearance of *Ptychadena* in the fossil record in the late Oligocene sediments (~25 mya) from the Nsungwe Formation in southwestern Tanzania.[40] This record represents the earliest of fossil anurans from Africa below the equator, and is the earliest definitive record of any family within the diverse ranoid clade Natatanura (sensu Blackburn et al.[41] and Frost et al.[42]). Other appearances of Ptychadenidae in the fossil record include those of the Middle Miocene at Beni Mellal in Morocco[43] and Pleistocene deposits from Madagascar[44].



prz, prezygapophysis; ns, neural spine; sc, sacral articular condyles; sd; sacral diapophysis; sf, spinal foramina

Scale bar equals 1 mm

Figure 2: (Top) *Ptychadena* indet. sp. 1 (Family: Ptychadenidae) from Langebaanweg, requisition number SAM-PQL-70839, in (a) dorsal, (b) ventral, (c) anterior and (d) posterior views. (Bottom) *Ptychadena* indet. sp. 2 (Family: Ptychadenidae) from Langebaanweg, requisition number SAM-PQL-71524, in (e) dorsal, (f) ventral, (g) anterior and (h) posterior views.

Phylogeography of Ptychadena

Extant ptychadenids comprise 55 species within the genera *Hildebrandtia, Lanzarana* and *Ptychadena*, and are widespread tropical and subtropical species distributed throughout Madagascar, sub-Saharan Africa, the Seychelles and the Mascarene islands[39], with the exception of two species found in the Nile Delta[40]. The distribution of the genus *Ptychadena* in southern Africa is delimited in areas where rainfall is above 60 mm in January and the minimum temperature is above 8 °C in September. These two variables represent 36.4% and 29.0% of the distribution of the genus in southern Africa, respectively (G.J.M. unpublished data). These two parameters describe the distribution of extant species from East London to the northeast, under the influence of the SRZ and increasing temperatures (Figure 1).

In the current WRZ, all modern species breed in winter, with the exception of two invasive species from the SRZ.[45] Some genera and families are found in both the WRZ and SRZ, and the remaining southern African families and genera are found in the SRZ and breed in summer (Table 1). Six amphibian families have been identified from LBW (Hyperoliidae, Brevicipitidae, Pxyicephalidae, Pipidae, Heleophrynidae and Bufonidae), and 19 different taxa, 10 of which remain unidentified, have been differentiated.[12] Genera identified include *Hyperolius, Breviceps, Tomopterna, Xenopus, Heleophryne* and *Amietophrynus.*[12] All the frog families and genera identified from LBW[12] currently occur in all three rainfall regimes (SRZ, WRZ and YRZ), with endemism to the WRZ only existing at species level in these genera[16]. Matthews et al.[12] tentatively identified *Kassina* (a genus restricted to the SRZ) – an identification which has subsequently been confirmed by computed tomography scans of comparative material, but the family to which it belongs (Hyperoliidae) contains other genera which are found in the current WRZ. Interestingly, histological studies done on the aquatic *Xenopus* (Family: Pipidae; common name: the African clawed frog) femora from LBW (paper in preparation) suggest marked seasonality as clear LAGS (lines of arrested growth) are evident.

Phylogeographic structuring in the distribution of anuran taxa with respect to the winter and summer rainfall regions has been recorded,[46] but little is known about the evolution of this pattern which clearly has deep roots. In the literature, the southern African amphibian fauna is commonly divided into two main groups – the so-called 'tropical' (a term used to describe the speciose tropical frog fauna distributed widely in the northeastern parts of southern Africa) and the unique 'Cape' frog faunas of the southwestern Cape which are characterised by extremely high rates of endemism and contain a large number of species with narrow niche envelopes.[16,17,47,48] These two groups correspond to the SRZ (the tropical frog fauna) and WRZ (the Cape frog fauna), and their breeding periods are delineated by mean annual precipitation in the summer or winter rainfall months, respectively.[17,48] The two rainfall regimes have such a high turnover in species that it is evident at a global scale.[49] Nonetheless, there are some genera that straddle both rainfall zones, but the population genetics of these taxa suggest deep genetic divisions, such that populations are confined to one zone or the other.[46]

Discussion

The palaeoenvironment at Langebaanweg

Fossil pollen is frequently used in palaeoenvironmental reconstruction to ascertain whether C_3 plants (those which fix and reduce inorganic CO_2 into organic compounds using the C_3 photosynthetic pathway) or C_4 plants (those which employ the C_4 photosynthetic pathway) dominated a palaeo-landscape. Grasses categorised as C_4 vegetation are typically associated with a summer rainfall regime, and C_3 grasses with the WRZ. Franz-Odendaal et al.'s[32] research on the large mammals from LBW indicated that the expected grazers (e.g. alcelaphine, hippopotamus and rhinoceros), as well as browsing species, showed δ13C values which indicated that LBW was a C_3 dominated environment. This finding was corroborated by a study of phytoliths from the site.[50] The small mammals from LBW, such as the rats and mice, suggest a fynbos component to the vegetation[9], as does fossil pollen[51,52]. The presence of cool-growing C_3 grasses at LBW during the deposition of MPPM sediments was taken to indicate that the present-day climatic regime of winter wet/summer dry was established early in the Pliocene epoch.[32] Using the vegetation to establish the rainfall regime is, however, questionable, given that fynbos and C_3 grasses both have a C_3 signature, and the contribution of C_3 grasses versus that of C_3 fynbos at LBW is undetermined. In addition, there are indications that the contribution from C_3 grasses to the general C_3 signature at LBW was small, as Stynder[10] suggests that fossil bovid species indicate that grass was scarce, and the environment may have been heavily wooded. The presence of woodland is supported by the presence of certain bird taxa, but arid and semi-arid landscapes are also represented.[7,9] The precise nature of the Early Pliocene vegetation at LBW remains elusive, has no modern analogue, and cannot contribute in any verifiable way to ascertaining the rainfall regime, or provide information on seasonality, at LBW. The fact that fynbos was present on the west coast during the Early Pliocene has been used as evidence of a WRZ; however, the premises on which this association is based are arguable and are discussed further in the next section.

Inception of the WRZ and aridification of the west coast

Of the eight frog genera found at LBW, six (*Hyperolius, Breviceps, Tomopterna, Heleophryne, Amietophrynus* and *Xenopus*) are found in summer, winter and year-round rainfall areas. The evidence of seasonality provided by *Kassina* is not clear-cut as, although this taxon is currently found in the SRZ, other members of the Hyperoliidae, such as *Semnodactylus*, are found in the current WRZ and YRZ. *Ptychadena* is, however, currently not found in the more southerly and western areas of South Africa, or in any parts of the WRZ[16], suggesting that this family has evolved and adapted to live in the SRZ.

The ptychadenids are a marker SRZ species, as they are found only in this rainfall zone, and the presence of *Ptychadena* at LBW thus provides strong evidence of a summer rainfall regime, or at least of a rainfall regime which included summer rainfall. The faunal evidence from LBW clearly indicates that at 5.1 mya the region was still receiving a relatively high rainfall, which fell partly, if not entirely, in summer.

This finding has important implications for studies involving the evolution of west coast flora and fauna, and the interpretation of phylogenetic studies which use molecular clock estimates to interpret the timing of lineage divergence. As mentioned previously, such studies have typically used the formation of the BUS as a proxy for the inception of the winter rainfall regime along the West Coast and the beginning of aridification.[6,24,32,33] Certain fynbos taxa such as Restionaceae, Ericaceae and Proteaceae formed part of the vegetation in southwestern Africa during the Palaeogene[53] and fossil pollen research indicates that fynbos taxa have formed part of west coast ecosystems since well before the inception of BUS and a winter rainfall regime[14,15,23,51,52,54]. Hoffman et al.'s[1] habitat reconstruction for the dated phylogenies of 12 plant clades from the Cape Floral Region in southern Africa indicated that they evolved in aseasonal rainfall environments. In addition, this research indicated that the initial development of aridity in the Cape Floral Region was not linked to the onset of strong rainfall seasonality – contradicting the common assumption that aridity and the development of a strongly seasonal winter rainfall regime on the west coast occurred at the same time and are effectively linked. This de-linking of these two variables indicates that commonly made assumptions in the literature are incorrect, and illustrates how much still remains to be elucidated about the evolution of the WRZ.

On a global scale, the middle Miocene is considered to have been a time of reduced seasonality and relatively stable climate with elevated temperatures and high humidity and rainfall.[1,6,52] Scisio et al.[52] studied the pollen from cores from the Miocene age Elandsfontyn Formation at LBW, which underlies the younger, main fossil-bearing deposits at the site. A strong contribution from plants that currently occur in tropical and subtropical conditions in the summer rainfall region in eastern South Africa is noted. The climate was wetter, and probably warmer, as indicated by the presence of palms and subtropical trees, the absence of Ericaceae, and the occurrence of only two Stoebe-type pollens. The fossil pollen evidence available which pre-dates LBW thus suggests that rainfall may have been either inter-annual, or characterised by a greater contribution of summer rain.[52]

Roberts et al.[15] suggested that the current pronounced aridity gradient from south to north on the west coast may have had its inception in the earlier Miocene. Hoffmann et al.[1] suggested that the Middle Miocene (13–17 mya) saw the development of perennial to weakly seasonal arid conditions, with the strongly seasonal rainfall regime of the west coast arising ~6.5–8 mya. Neumann and Bamford[6] suggest that aridification of the west coast began post Middle Miocene. Based on a palynology study of sediments retrieved from the Atlantic off the mouth of the Orange River, Du pont et al.[54] note that between 10 mya and 6 mya, pollen types from plants of tropical affinity disappeared, and those from the Cape flora gradually increased. The onset of aridity on the west coast has thus been postulated to pre-date the LBW fossils by 3–5 mya. This postulate has, however, been based on palaeoclimatic proxies situated much further to the north of LBW in Namibia[31,55] and the Orange River[54]. The LBW frog community indicates a relatively high rainfall[12], as do some other fossil taxa[13], and the fact that woodland appears to have been widespread[10]. The evidence from LBW strongly contradicts assertions that aridification of the west coast in the region was well established by 5.1 mya, although such a scenario was certainly applicable further northwards, with aridity in the Namib dating back to 16 Ma.[55] However, relatively moister conditions appear to have prevailed in the southern, as opposed to the central, Namib desert during the Miocene and the flora appears to have remained C_3 in both regions until the Pliocene.[56]

The marine evidence mentioned previously is inconclusive regarding the effects of fluctuations in the BUS and SSTs on the rainfall regime during the Early Pliocene. Much uncertainty therefore exists as to the exact status of the strength, and effect, of the BUS and SSTs, and the inferences drawn from other palaeoclimatic proxies are contestable. Neither the fossil evidence, the fossil pollen data, nor the fact that C_3 plants predominate, provide direct evidence of winter rainfall at LBW, or indeed of the period of inception of the WRZ.

Conclusions

The majority of anuran taxa at LBW are currently found in the SRZ and WRZ, and consequently are not diagnostic in terms of the rainfall regime. However, the identification of two *Ptychadena* species from LBW, as well as the presence of *Kassina*, provides new evidence of a summer rainfall regime, or of at least some significant summer rainfall, at ~5.1 mya on the southwestern coast of southern Africa. This evidence is a significant contribution to our understanding of the evolution of seasonality on the west coast in the LBW region and challenges the commonly held assumption that the WRZ has been established on the west coast since the later Miocene/Early Pliocene. The presence of a *Ptychadena* species also contributes to our knowledge of the palaeo-distribution of the family, which is currently widespread in sub-Saharan Africa, and has endured since the Oligocene in Tanzania.

Acknowledgements

The support of the DST-NRF Centre of Excellence in Palaeosciences is hereby acknowledged. G.J.M. thanks the DST-NRF Centre of Excellence for Invasion Biology for their support. D.L.R. acknowledges support from the National Research Foundation and the Department of Geography, Faculty of Natural and Agricultural Sciences at the University of the Free State.

This paper is dedicated to the scientific contributions of Dr Dave Roberts, our co-author, who passed away during the final preparation of this manuscript. His contributions as a geologist were so wide-ranging that they spanned the interests of a great many scientific disciplines, of which this work is one such example. His insights and enthusiasm will be greatly missed.

Authors' contributions

T.M. was the project leader; undertook the fossil research and provided the first draft of the manuscript. D.L.R. provided information on climatic change and conditions over the Miocene and Pliocene and made conceptual contributions. G.J.M. made conceptual contributions and provided information on the frog taxa mentioned in the paper. All three authors contributed to writing the manuscript.

References

1. Hoffmann V, Verboom GA, Cotterill FPD. Dated plant phylogenies resolve neogene climate and landscape evolution in the Cape Floristic Region. PLoS One. 2015;10(9), e0137847, 25 pages. Available from: http://dx.doi.org/10.1371/journal.pone.0137847
2. Meadows ME, Baxter AJ. Late Quaternary palaeoenvironments of the south-western Cape, South Africa: A regional synthesis. Quat Int. 1999;57/58:193–206. http://dx.doi.org/10.1016/S1040-6182(98)00060-3
3. Chase BM, Meadows ME. Late Quaternary dynamics of southern Africa's winter rainfall zone. Earth Sci Rev. 2007;84:103–138. http://dx.doi.org/10.1016/j.earscirev.2007.06.002
4. Chase BM, Boom A, Carr AS, Meadows ME, Reimer PJ. Holocene climatic change in southernmost South Africa: Rock hyrax middens record shifts in the southern westerlies. Quat Sci Rev. 2013;82:199–205. http://dx.doi.org/10.1016/j.quascirev.2013.10.018
5. Chase BM, Lim S, Chevalier M, Boom A, Carr AS, Meadows ME, et al. Influence of tropical easterlies in southern Africa's winter rainfall zone during the Holocene. Quat Sci Rev. 2015;107:138–148. http://dx.doi.org/10.1016/j.quascirev.2014.10.011
6. Neumann FH, Bamford MK. Shaping of modern southern African biomes: Neogene vegetation and climate changes. Trans Roy Soc S Afr. 2015;70(3):195–212. http://dx.doi.org/10.1080/0035919X.2015.1072859
7. Manegold AA, Louchart JC, Elaznowski A. The Early Pliocene avifauna of Langebaanweg (South Africa): A review and update. In: Gohlich UB, Koh A, editors. Paleornithological Research, Proceedings of the 8th International Meeting of the Society of Avian Paleontology and Evolution. Vienna: Natural History Museum; 2013. p. 135–152.
8. Matthews T. A review of the Muridae of Langebaanweg (Mio-Pliocene, West Coast, South Africa). Published abstracts of the Langebaanweg 2006: Mini-symposium and workshop. Afr Nat Hist. 2006;2:188–189.
9. Matthews T, Parkington JE, Denys C. Community evolution of Neogene micromammals from Langebaanweg 'E' Quarry and other west coast fossil sites, south-western Cape, South Africa. Palaeogeogr Palaeocl. 2007;245:332–352. http://dx.doi.org/10.1016/j.palaeo.2006.08.015
10. Stynder DD. Fossil bovid diets indicate a scarcity of grass in the Langebaanweg 'E' Quarry (South Africa) late Miocene/early Pliocene environment. Paleobiology. 2011;37(1):126–139. http://dx.doi.org/10.1666/09074.1
11. Diekmann B, Falker M, Kuhn G. Environmental history of the south-eastern South Atlantic since the Middle Miocene: Evidence from the sedimentological records of ODP Sites 1088 and 1092. Sedimentology. 2003;50:511–529. http://dx.doi.org/10.1046/j.1365-3091.2003.00562.x
12. Matthews T, Van Dijk E, Roberts DL, Smith RMH. An early Pliocene (51 Mya) fossil frog community from Langebaanweg, south-western Cape, South Africa Afr J Herpetol. 2015;64(1):39–53. http://dx.doi.org/10.1080/21564574.2014.985261
13. Hendey QB. Cenozoic geology and palaeoecology of the fynbos region. In: Deacon HJ, Hendey QB, Lambrechts JJN, editors. Fynbos palaeoecology: A preliminary synthesis. South African National Scientific Programmes Report no 75. Cape Town: Mills Litho; 1983. p. 35–60.
14. Roberts DL, Matthews T, Herries AIR, Boulter C, Smith RMH, Dondo C, et al. Regional and global context of the Late Cenozoic Langebaanweg (LBW) palaeontological site: West Coast of South Africa. Earth Sci Rev. 2011;106(3):191–310. http://dx.doi.org/10.1016/j.earscirev.2011.02.002
15. Roberts DL, Scisio L, Herries AIR, Scott L, Bamford MK, Musekiwa C, et al. Miocene fluvial systems and palynofloras at the southwestern tip of Africa: Implications for regional and global fluctuations in climate and ecosystems. Earth Sci Rev. 2013;124:184–201. http://dx.doi.org/10.1016/j.earscirev.2013.05.001
16. Minter LR, Burger M, Harrison JA, Braak HH, Bishop PJ, Kloepfer D. Atlas and red data book of the frogs of South Africa, Lesotho and Swaziland SI/MAB Series no 9. Washington DC: Smithsonian Institution; 2004.
17. Schreiner C, Rodder D, Measey GJ. Using models to test Poynton's predictions. Afr J Herpetol. 2013;62:49–62. http://dx.doi.org/10.1080/21564574.2013.794865
18. Colville JC, Potts AJ, Bradshaw PL, Measey GJ, Snijman D, Picker MD, et al. Floristic and faunal Cape biochoria: Do they exist? In: Allsopp N, Colville JF, Verboom GA, editors. Fynbos: Ecology, evolution and conservation of a megadiverse region. Cape Town: Oxford University Press; 2014. p. 73–93. http://dx.doi.org/10.1093/acprof:oso/9780199679584.003.0004

19. Tyson PD. Climatic change and variability in southern Africa. Cape Town: Oxford University Press; 1986.

20. Chase BM, Thomas DSG. Multiphase late Quaternary aeolian sediment accumulation in western South Africa: Timing and relationship to palaeoclimatic changes inferred from the marine record. Quat Int. 2007;166(1):29–41. http://dx.doi.org/10.1016/j.quaint.2006.12.005

21. Diester-Haass L, Meyers PA, Vidal L. The late Miocene onset of high productivity in the Benguela Current upwelling system as part of a global pattern. Marine Geol. 2002;180:87–103. http://dx.doi.org/10.1016/S0025-3227(01)00207-9

22. Coetzee JA. Late Cainozoic palaeoenvironments of southern Africa In: Van Zinderen Bakker Sr EM, editor. Antarctic glacial history and world palaeoenvironments. Rotterdam: AA Balkema; 1978. p. 25–32. http://dx.doi.org/10.1017/S0032247400001819

23. Coetzee JA, Rogers J. Palynological and lithological evidence for the Miocene palaeoenvironment in the Saldhana region (South Africa). Palaeogeogr Palaeocl. 1982;39:71–85. http://dx.doi.org/10.1016/0031-0182(82)90073-6

24. Siesser WG. Aridification of the Namib Desert: Evidence from oceanic cores. Van Zinderen Bakker Sr EM, editor. Antarctic glacial history and world palaeoenvironments. Rotterdam: AA Balkema; 1978. p. 105–113. http://dx.doi.org/10.1017/S0032247400001819

25. Gasse F. Hydrological changes in the African tropics since the Last Glacial Maximum. Quat Sci Rev. 2000;19:189–211. http://dx.doi.org/10.1016/S0277-3791(99)00061-X

26. Farmer EC, deMenocal PB, Marchitto TM. Holocene and deglacial ocean temperature variability in the Benguela upwelling region: Implications for low latitude atmospheric circulation. Paleoceanography. 2005;20, PA2018. http://dx.doi.org/10.1029/2004PA001049

27. Luyt J, Lee-Thorp J, Avery G. New light on middle Pleistocene west coast environments from Elandsfontein, western Cape, South Africa. S Afr J Sci. 2000;96:399–403.

28. Heinrich S, Zonneveld KAF, Bickert T, Willems H. The Benguela upwelling related to the Miocene cooling events and the development of the Antarctic Circumpolar Current: Evidence from calcareous dinoflagellate cysts. Paleoceanography. 2011;26, PA3209. http://dx.doi.org/10.1029/2010PA002065

29. Etourneau J. Tectonically driven upwelling. Nat Geosci. 2014;7:698–699. http://dx.doi.org/10.1038/ngeo2258

30. Jung G, Prange M, Schultz M. Uplift of Africa as a potential cause for Neogene intensification of the Benguela upwelling system. Nat Geosci. 2014;7:741–747. http://dx.doi.org/10.1038/ngeo2249

31. Roberts DL, Brink JS. Dating and correlation of Neogene coastal deposits in the Western Cape (South Africa): Implications for Neotectonism. S Afr J Geol. 2002;105(4):337–352. http://dx.doi.org/10.2113/1050337

32. Franz-Odendaal T, Lee-Thorp JA, Chinsamy A. New evidence for the lack of C_4 grassland expansions during the Early Pliocene at Langebaanweg, South Africa. Paleobiology. 2002;28(3):378–388. http://dx.doi.org/10.1666/0094-8373(2002)028<0378:NEFTLO>2.0.CO;2

33. Altwegg R, West A, Gillson L, Midgley GF. Impacts of climate change in the Greater Cape Floristic region. In: Allsopp N, Colville JF, Verboom GA, editors. Fynbos: Ecology, evolution and conservation of a megadiverse region. Cape Town: Oxford University Press; 2014. p. 299–320. http://dx.doi.org/10.1093/acprof:oso/9780199679584.003.0013

34. Rommerskirchen F, Condon T, Mollenhauer G, Dupont L, Schefuss L. Miocene to Pliocene development of surface and sub-surface temperatures in the Benguela Current system. Paleoceanography. 2011;26, PA3216. Available from: http://dx.doi.org/10.1029/2010PA002074

35. Pether J. Late Tertiary and early Quaternary marine deposits of the Namaqualand coast, Cape Province: New perspectives. S Afr J Sci. 1986;82:464–470.

36. Pether J. Molluscan evidence for enhanced deglacial advection of Agulhas water in the Benguela Current, off south-western Africa. Palaeogeogr Palaeocl. 1994;111:99–117. http://dx.doi.org/10.1016/0031-0182(94)90350-6

37. Kensley B. Pliocene marine invertebrates from Langebaanweg, Cape Province. Ann S Afr Mus. 1972;60:173–190.

38. Hendey QB. The Pliocene occurrences in 'E' Quarry, Langebaanweg, South Africa. Ann S Afr Mus. 1976;69: 215–247.

39. Frost DR. Amphibian species of the world: An online reference, Version 60 [database on the Internet]. No date [cited 2015 Sep 11]. Available from: http://researchamnhorg/herpetology/amphibia/indexhtml

40. Blackburn DC, Wake DB. Class Amphibia Gray, 1825. In: Zhang ZQ, editor. Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Auckland: Magnolia Press; 2011. p. 39–55.

41. Blackburn DC, Roberts EM, Stevens NJ. The earliest record of the endemic African frog family Ptychadenidae from the Oligocene Nsungwe Formation of Tanzania. J Vert Pal. 2015;35(2), Art. #e907174, 5 pages. http://dx.doi.org/10.1080/02724634.2014.907174

42. Frost DR, Grant T, Faivovich J, Bain RH, Haas A, Haddad CFB, et al. The amphibian tree of life. Bull Am Mus Nat Hist. 2006;297:1–370. http://dx.doi.org/10.1206/0003-0090(2006)297[0001:TATOL]2.0.CO;2

43. Vergnaud-Grazzini C. Les amphibiens du Miocene de Beni-Mellal [Amphibians of the Miocene from Beni-Mellal]. Notes et Mémoires du Service Géologique du Maroc. 1966;27:43–74. French.

44. Burney DA, Vasey N, Godfrey LR, Ramilisonina WL, Jungers M, Ramarolahy M, et al. New findings at Andrahomana Cave, southeastern Madagascar. J Cave Karst Stud. 2008;70:13–24.

45. Measey GJ, Davies S. Struggling against domestic exotics at the southern end of Africa. Froglog. 2011;97:28–30.

46. Tolley KA, Bowie RCK, Measey GJ, Price BW, Forest F. The shifting landscape of genes since the Pliocene: Terrestrial phylogeography in the Greater Cape Floristic Region. In: Allsopp N, Colville JF, Verboom GA, editors. Fynbos: Ecology, evolution and conservation of a megadiverse region. Cape Town: Oxford University Press; 2014. p. 142–163. http://dx.doi.org/10.1093/acprof:oso/9780199679584.003.0007

47. Poynton JC. Afrotemperate amphibians in southern and eastern Africa: A critical review. Afr J Herpetol. 2013;62(1):5–20. http://dx.doi.org/10.1080/21564574.2013.794866

48. Mokhatla MM, Rödder D, Measey GJ. Assessing the effects of climate change on distributions of Cape Floristic Region amphibians. S Afr J Sci. 2015;111(11–12), Art. #2014-0389, 7 pages. http://dx.doi.org/10.17159/sajs.2015/20140389

49. Holt BG, Lessard JP, Borregaard MK, Fritz SA, Araújo MB, Dimitrov D, et al. An update of Wallace's zoogeographic regions of the world. Science. 2013;339(6115):74–78. http://dx.doi.org/10.1126/science.1228282

50. Rossouw L, Stynder D, Haarhoff P. Evidence for opal phytolith preservation in the Langebaanweg 'E' Quarry Varswater Formation and its potential for paleohabitat reconstruction. S Afr J Sci. 2009;105:223–227.

51. Scott L. Pollen evidence for vegetational and climatic change in southern Africa during the Neogene and Quaternary. In: Vrba ES, Denton GH, Partridge TC, Burckle LH, editors. Paleoclimate and evolution with emphasis on human origins Yale: Yale University Press; 1995. p. 65–76.

52. Scisio L, Neumann FH, Roberts DL, Tsikos H, Scott L, Bamford MK. Fluctuations in Miocene climate and sea levels along the south-western South African coast: Inferences from biogeochemistry, palynology and sedimentology. Pal Afr. 2013;48:2–18.

53. Linder HP. The radiation of the Cape flora, southern Africa. Biol Rev. 2003;78(4):597–638. http://dx.doi.org/10.1017/S1464793103006171

54. Dupont LM, Linder HP, Rommerskirchen F, Schefuß E Climate-driven rampant speciation of the Cape flora. J Biogeog. 2011;38(6):1059–1068. http://dx.doi.org/10.1111/j.1365-2699.2011.02476.x

55. Senut B, Pickford M, Ségalen L. Neogene desertification of Africa. Compt Rend Geosci. 2009;341:591–602. http://dx.doi.org/10.1016/j.crte.2009.03.008

56. Senut B, Ségalen L Neogene palaeoenvironments of the Namib Desert: A brief synthesis. Trans Royal Soc S Afr. 2014;69:205–211. Available from: http://dx.doi.org/10.1080/0035919X.2014.958605

Differential involvement of ascorbate and guaiacol peroxidases in soybean drought resistance

AUTHORS:
Makoena J. Moloi[1]
Obed J. Mwenye[2]
Rouxlene van der Merwe[1]

AFFILIATIONS:
[1]Department of Plant Sciences, University of the Free State, Bloemfontein, South Africa
[2]Bvumbwe Research Station, Limbe, Malawi

CORRESPONDENCE TO:
Makoena Moloi

EMAIL:
makoena.moloi@gmail.com

DATES:
Received: 26 Jan. 2016
Revised: 30 Mar. 2016
Accepted: 23 May 2016

KEYWORDS:
antioxidative enzymes; drought stress; maturity stage; drought tolerance

HOW TO CITE:
Moloi MJ, Mwenye OJ, Van der Merwe R. Differential involvement of ascorbate and guaiacol peroxidases in soybean drought resistance. S Afr J Sci. 2016;112(9/10), Art. #2016-0028, 4 pages. http://dx.doi.org/10.17159/sajs.2016/20160028

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa)

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

Soybean (*Glycine max* L.) is a small but growing component of the agricultural economy of South Africa and is predicted to become a major crop in Africa because of its high protein content. Drought induction at flowering or early stages of pod development has detrimental effects on soybean yield. As antioxidative enzymes play a protective role in plants during various abiotic stress conditions, this study was conducted to investigate how ascorbate (Enzyme Commission (EC) number 1.11.1.1) and guaiacol (EC: 1.11.1.7) peroxidases are involved in soybean drought resistance at different maturity stages (flowering and pod development). We also investigated whether the levels of these enzymes decline with plant maturity. Three tolerant soybean genotypes (G1, G2, G3) and a susceptible genotype (G4*) were used. These cultivars were categorised according to their sensitivity to drought stress in previous studies. The activity of ascorbate peroxidase was significantly induced by drought stress at both growth stages with higher activity in the resistant than susceptible plants, strongly supporting the protective role of this enzyme against drought stress at both developmental stages. The guaiacol peroxidase activity was induced to higher levels in the resistant than in the susceptible plants at flowering only, with no significant increase observed at pod development stage, indicating its selective protective involvement against drought stress. Interestingly, the levels of these enzyme activities were induced in all cultivars at both developmental stages, irrespective of drought stress, indicating that their activities increased with maturity.

Significance:

- Guaiacol peroxidase is selectively involved in soybean drought resistance at flowering stage.
- The upregulation of ascorbate peroxidase activity at both growth stages in drought-resistant cultivars suggests that this enzyme could be used as a biochemical marker of drought resistance in soybeans.
- In contrast to the literature, activities of both enzymes increased with maturity irrespective of whether the plant is drought susceptible or resistant.

Introduction

Soybean (*Glycine max* L.) is a small but growing component of the agricultural economy of South Africa[1] and is predicted to become a major crop in Africa because of its high protein content[2]. Among important abiotic factors that influence soybean growth and yields are temperatures above 30 °C and lower than 13 °C for long periods during the flowering stage (which inhibits flower and seed development) as well as drought stress imposed during flowering and pod formation stages.[3] Drought stress occurs when available water in the soil is reduced and atmospheric conditions cause continuous loss of water through transpiration or evaporation. In most crops, inhibition of growth and yield are mainly associated with altered metabolic functions such as reduced photosynthesis.[4]

Plants respond to drought stress with a cascade of biochemical reactions such as production of abscisic acid, which is aimed at facilitating stomatal closure thereby reducing water loss through transpiration. Although this action reduces water loss, it limits carbon dioxide fixation and reduces regeneration of $NADP^+$ by Calvin cycle, which results in increased formation of oxygen radicals or reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2), superoxide anion (O_2^-) and hydroxyl radicals (OH^-) through enhanced leakage of electrons to molecular oxygen.[5] Studies show that ROS such as H_2O_2 can act as signal molecules for induction of defence responses in plants when moderately produced.[6,7] However, excessive production of these molecules may lead to oxidative stress, which in turn may have damaging effects on the photosynthetic pigments, membrane lipids, proteins and nucleic acids.[8] To avoid oxidative stress and to thrive, plants need to keep ROS production in the cells to a minimum,[9] which can be achieved through employment of different antioxidative mechanisms that may be enzymatic or non-enzymatic in nature. The major ROS scavenging mechanisms include enzymes of the ascorbate–glutathione pathway (e.g. ascorbate peroxidase (APX) and glutathione reductase), guaiacol peroxidase (POD) and catalase.[10] These enzymes may act independently or in conjunction to catalyse conversion of H_2O_2 to water and O_2,[11,12] thereby minimising damage within the plant.

However, it is important to note that ROS production in plants depends on the stress intensity, plant species and genotype as well as the developmental stage.[13] Drought tolerance was found to be correlated with the ROS scavenging capability.[14] However, the degree to which the activities of antioxidative enzymes are elevated or inhibited under drought stress is variable among plant species and even between cultivars of the same species.[15] As a consequence of the increase in intensity of drought spells in South Africa, we undertook this study to investigate whether APX and POD are differentially or similarly involved in soybean drought resistance at the two maturity stages (flowering and pod development). We also sought to clarify whether the levels of enzyme activities were affected by the different growth stages, as numerous studies have reported decreases or unchanged enzyme activities with maturity. We therefore report, for the first time, on the involvement of APX and POD when drought stress is applied at flowering and pod development. Results will further indicate whether these enzymes may be used as biochemical markers of drought resistance in soybeans.

Materials and methods

Drought tolerant (G1, G2, G3) and susceptible (G4*) soybean cultivars were planted in lysimeter units (1.40 m long), each filled with air-dried A (0.25 m, top) and B (1.05 m, bottom) horizon of Bainsvlei Amalia 2300 soil with filter sand at the bottom of the cylinders and a mulch of styrofoam at the top of the cylinders. The choice of cultivars was based on previous glasshouse and field trials, which showed that G1, G2 and G3 were less sensitive whereas G4* was very sensitive to drought stress. Although the G1, G2 and G3 cultivars were all tolerant, there were differences in their mode of drought resistance (results not shown).

The experiment was conducted in replicates of three per cultivar per treatment, in the glasshouse under controlled temperature (25 °C day and 18 °C night) in a randomised complete block design. Drought stress was applied according to literature[16,17] at the beginning of flowering and pod development stages by withdrawing water supply to the plants for 14 days, followed by optimal watering (optimal watering refers to lysimeters being watered every day based on the amount of depleted soil water and irrigated back to field capacity). Other sets of tolerant and susceptible plants, representing the controls, received optimal water supply throughout the experimental period. At the end of each stress treatment, the top three fully expanded leaves were harvested on the stressed as well as the control plants and immediately frozen in liquid nitrogen before storage at -20 °C for the antioxidative enzyme analysis. Two plants per treatment were sampled. The obtained data were subjected to analysis of variance using GenStat Release 17.1 software to separate the sources of variation.

Enzyme extract preparation and assay

Enzyme extracts were prepared in accordance with Pukacka and Ratajczak[18]. Pre-weighed leaves (0.5 g) of each treatment were homogenised to a fine paste on ice using a mortar and pestle in 5 mL 50 mM potassium phosphate buffer (pH 7.0) containing 1 mM EDTA, 2% PVPP, 0.1% Triton X-100 and 1 mM ascorbate. The homogenate was centrifuged at 15 000 x *g* for 20 min at 4 °C. The resulting aliquot was used as the enzyme extract. Three replicates per assay were used.

The APX assay was performed according to a method described by Mishra et al.[19] with modifications. The reaction mixture (1 mL) consisted of 500 μL 50 mM phosphate buffer (pH 7.0), 200 μL 0.1 mM H_2O_2, 150 μL 0.5 mM sodium ascorbate, 50 μL 0.1 mM EDTA and 100 μL enzyme. A decrease in absorbance as a result of ascorbate oxidation was measured at 290 nm for 5 min at 20 °C against a blank in which the enzyme was replaced with phosphate buffer. An extinction coefficient of 2.8 $mM^{-1}cm^{-1}$ was used to calculate enzyme activity.

A modified method described by Zieslin and Ben-Zaken[20] was used for determination of POD. The reaction mixture contained 50 μL 0.2 M H_2O_2, 100 μL 50 mM guaiacol, 340 μL distilled H_2O, 500 μL 80 mM phosphate buffer (pH 5.5) and 10 μL enzyme. An increase in absorbance as a result of tetraguaiacol formation was measured at 470 nm for 3 min at 30 °C. The blank contained all reagents except for the enzyme, which was replaced with phosphate buffer. An extinction coefficient of 26.6 $mM^{-1}cm^{-1}$ was used to calculate enzyme activity. Protein concentration determination was done according to Bradford[21].

Results

Highly significant differences in APX ($p \leq 0.001$) were observed between drought-stressed and well-watered soybean genotypes at different growth stages (flowering and pod development). In contrast, a significant effect of drought stress on POD ($p \leq 0.01$) was observed only at flowering and not at pod development stage. Activities of both APX and POD were highly significant for all cultivars, which indicated that there were large differences in antioxidative enzyme levels across the four genotypes. This finding was applicable for both flowering and pod development stages. Water treatment by genotype interaction was highly significant at flowering stage ($p \leq 0.001$) for both enzymes. However, at pod development stage of all studied cultivars, drought stress had a selective significant effect ($p \leq 0.01$) on APX but not on POD (Table 1).

Table 1: Mean square values of ascorbate peroxidase (APX) and guaiacol peroxidase (POD) enzyme activities measured on four soybean cultivars for two water treatments at flowering and pod development stages

Source of variation	Flowering		Pod development	
	APX	POD	APX	POD
Replication	0.0000014	0.000560	0.0000780	0.01978
Water treatment	0.00009691***	0.0825323***	0.0011925***	0.04691
Genotype	0.00002763***	0.0072832***	0.0031369***	0.55354***
Water treatment x genotype	0.0001158***	0.0076786***	0.0008940***	0.06976
Residual	0.0000039	0.00057	0.0001800	0.04639
Coefficient of variation%	11.9000	10.2000	12.5000	15.7000

****p*≤0.001*

At flowering, drought stress induced significant increases in APX activities of G1 (2.5-fold), G2 (1.4-fold) and G3 (1.2-fold) cultivars, whereas in the susceptible cultivar (G4*), drought stress led to a substantial decrease (43%) in APX activity. Similarly at pod developmental stage, drought stress led to substantial increases in APX activities of G1 (1.3-fold), G2 (1.4-fold) and G3 (1.1-fold) cultivars and a decrease (15%) in the susceptible cultivar (Figure 1).

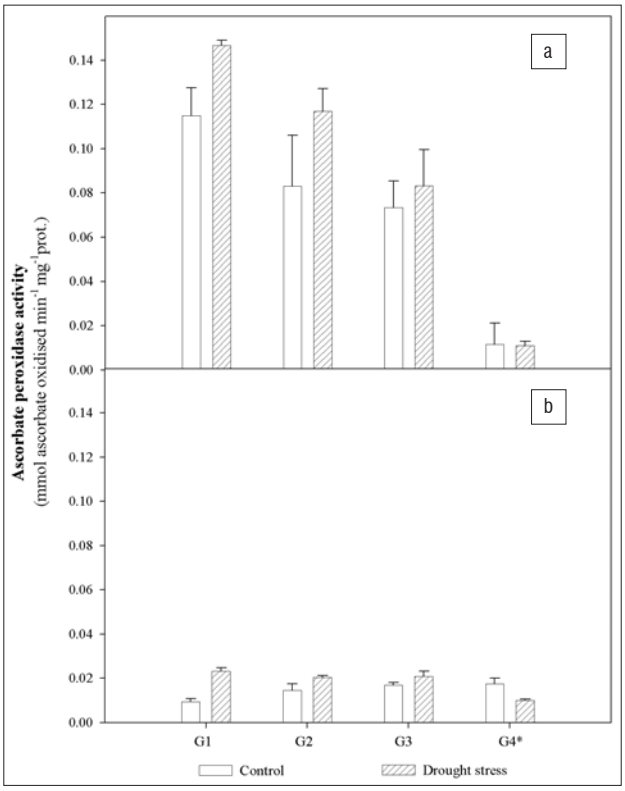


Figure 1: The effect of drought stress on the ascorbate peroxidase activity of tolerant (G1, G2, G3) and susceptible (G4*) soybean plants at (a) pod development and (b) flowering stages. Values are means±s.d. (n=3).

At flowering, resistant cultivars (G1, G2, G3) responded to drought stress with a significant increase in POD activity of 2.2-fold, 2.0-fold and 1.4-fold, respectively. In the susceptible cultivar at the same developmental stage, no significant increase in POD activity post drought stress

application was observed. In contrast to the APX activity results and flowering stage above, there were no significant changes in POD activity of the various cultivars at pod development stage (Figure 2).

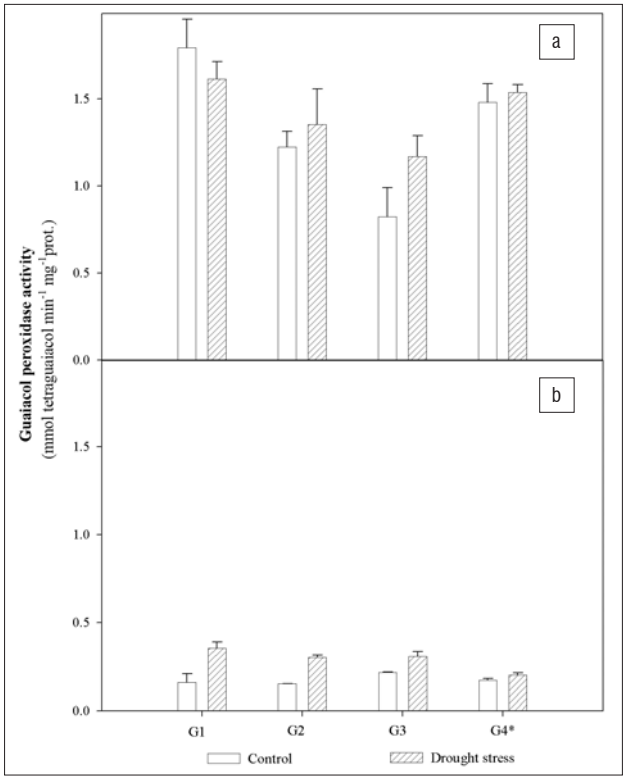


Figure 2: The effect of drought stress on the guaiacol peroxidase activity of tolerant (G1, G2, G3) and susceptible (G4*) soybean plants at (a) pod development and (b) flowering stages. Values are means±s.d. (n=3).

Discussion

Drought at flowering or early pod development stages significantly increases the rate of pod abortion and consequently decreases seed yield.[22] Prolonged drought stress leads to overproduction of ROS[23]; therefore plants need to respond with a battery of antioxidative enzymes in order to thrive[9]. APX was significantly affected when drought stress was applied at flowering and pod development stages (Table 1), confirming that this enzyme is highly sensitive to drought stress because of its high affinity for H_2O_2.[24] Further evidence showed that APX was significantly different among the cultivars used in this study (Table 1), showing that sensitivity to drought stress for this enzyme cannot be generalised. This finding further confirmed earlier sensitivity trials (results not available) that showed that the mode of tolerance for these cultivars was different. The measure of various antioxidative enzyme activities and/or expression may be used as an approach to assess the involvement of a scavenging system during drought stress.[15] Numerous studies that have been conducted in various plants are in support of these findings. Drought stress also induced increases in many antioxidative enzymes including APX in sunflowers.[25] Likewise in soybean (cultivar Enrei), drought stress led to a substantial increase in APX activity, measured using Western blot, enzyme activity assay and biophoton emission.[26] To show that the response of plants to drought stress is species or cultivar specific, a different response was observed in mycorrhizal soybeans in which APX activity declined significantly after drought stress introduction.[27] Additionally, drought stress had no effect on APX transcript levels in spinach[28], whereas it led to a significant decrease in this enzyme's activity in tomato[29].

Although APX activity was significantly increased by drought stress at both developmental stages in resistant cultivars, evidence also showed that the levels of activity at these stages were not the same. Activity levels at pod development were higher than at flowering (Figure 1), suggesting that activity of this enzyme intensified with maturity, in contrast to numerous studies that have reported a decrease. Antioxidative enzyme activity of barley under cadmium stress remained the same in two maturity stages.[30] Mature leaves of neem and pigeon pea had lower APX activity than younger leaves.[31] Antioxidative enzyme activities declined significantly with progressive growth stages in drought-stressed maize.[32] Results further indicated that APX activity in all tolerant plants (G1, G2, G3) was significantly upregulated by drought stress irrespective of the growth stage. In the susceptible cultivar (G4*), however, the enzyme was downregulated (Figure 1). As APX is one of the enzymes responsible for scavenging H_2O_2 which may be produced during water deficits,[23] significant increases in activity in the tolerant cultivars and a substantial decline in the susceptible cultivar suggest that tolerance to drought stress for these cultivars was somehow associated with this enzyme. These results indicate that APX could be used as a biochemical marker for drought resistance in soybean, as did those of Zarei et al.[33] who showed that APX activity induction during drought stress was correlated with drought tolerance in tobacco.

During stress conditions, PODs are often the first antioxidative enzyme activities to be altered.[34] Similarly to APX, lower POD activity at flowering than at pod development showed that induction of antioxidative enzymes increased with maturity in these cultivars. Significantly higher increases in POD activity of tolerant than of susceptible cultivars after drought stress induction at flowering (Figure 2) further showed that this enzyme was part of the drought resistance mechanisms when stress was introduced at flowering. Lack of significant interaction between water treatment (i.e. drought stress) and genotypes at pod development stage for POD (Table 1) showed that POD was an essential part of the ROS detoxification system at this stage for the majority of cultivars. However, this may not mean that these cultivars were not tolerant to drought stress; other studies have shown that if any of the antioxidative enzymes are upregulated under water-stress conditions, the genotype is still likely to be tolerant.[34] Studies in *Brassica napus* (rapeseed), *Helianthus annuus* (sunflowers) and *Triticum aestivum* (wheat), support the finding that upregulation of POD is associated with drought tolerance.[13,25,34,35]

In conclusion, the upregulation of APX and POD activities in tolerant soybean cultivars suggests their involvement in drought stress resistance. However, APX and POD were found to be differentially involved, with APX being part of the drought resistance mechanisms at both maturity stages whereas POD was involved at flowering only. The presented evidence shows that APX could be used as a biochemical marker for drought resistance in soybeans. Interestingly, the levels of these enzyme activities were induced in all cultivars at both developmental stages, irrespective of drought stress, indicating that their activities increased with maturity.

Acknowledgement

We acknowledge financial support from the National Research Foundation (South Africa).

Authors' contributions

M.J.M. wrote the manuscript, performed the experiments and analysed the data; O.J.M. performed the glasshouse trials; and R.v.d.M. was the leader of the project and acquired the funding.

References

1. De Beer A, Prinsloo T. The national soybean cultivar trials in South Africa – 34 years experiences and progress. Potchefstroom: ARC-GCI; 2013. Available from: http://www.proteinresearch.net/html_images/wsrc2013/18-february-session-2/134_debeer-as.pdf
2. Sinclair TR, Marrou H, Soltani A, Vadez V, Chandolu KC. Soybean production potential in Africa. Glob Food Sec. 2014;3:31–40. http://dx.doi.org/10.1016/j.gfs.2013.12.001
3. Pannar Seed (Pty) Ltd. Soybeans production guide. Production guide series: Soybeans. Greytown: Pannar Seed; 2006. p. 1–15.
4. Jaleel CA, Manivannan P, Wahid A, Farooq M, Al-Juburi HJ, Somasundaram R, et al. Drought stress in plants: A review on morphological characteristics and pigments composition. Int J Agric Biol. 2009;11:100–105.

5. Arora A, Sairam RK, Sriuastava GC. Oxidative stress and antioxidative system in plants. Curr Sci. 2002;82:1227–1238.

6. Chamnongpol S, Willekens H, Moeder W, Langebartels C, Sandermann H, Van Montagu M, et al. Defence activation and enhance pathogen tolerance induced by H_2O_2 in transgenic plants. Proc Natl Acad Sci USA. 1998;58:5818–5823.

7. Moloi MJ, Van der Westhuizen AJ. The reactive oxygen species are involved in the resistance responses of wheat to the Russian wheat aphid. J Plant Physiol. 2006;163:1118–1125. http://dx.doi.org/10.1016/j.jplph.2005.07.014

8. Mittler R. Oxidative stress, antioxidants and stress tolerance. Trends Plant Sci. 2002;7:405–410. http://dx.doi.org/10.1016/S1360-1385(02)02312-9

9. Del Rio LA, Corpas CJ, Sandalio LM, Palma JM, Gomez M, Barroso JB. Reactive oxygen species, antioxidant systems, and nitric oxide in peroxisomes. J Exp Bot. 2002;53:1255–1272. http://dx.doi.org/10.1093/jexbot/53.372.1255

10. Xu PL, Guo YK, Bai JG, Shang L, Wang XJ. Effects of long-term chilling on ultrastructure and antioxidative activity in leaves of two cucumber cultivars under low light. Physiol Plant. 2008;132:467–478. http://dx.doi.org/10.1111/j.1399-3054.2007.01036.x

11. Noctor G, Foyer CH. Ascorbate and glutathione: Keeping active oxygen under control. Annu Rev Plant Phys. 1998;49:249–279. http://dx.doi.org/10.1146/annurev.arplant.49.1.249

12. Gratao PL, Polle A, Lea PJ, Azevedo RA. Making the life of heavy metal-stressed plants a little easier. Funct Plant Biol. 2005;32:481–494. http://dx.doi.org/10.1071/FP05016

13. Abedi T, Pakniyat H. Antioxidant enzyme changes in response to drought stress in ten cultivars of oilseed rape (*Brassica napus* L.). Czech J Genet Plant Breed. 2010;46:27–34.

14. Selote DS, Khanna-Chopra. Drought-induced spikelet sterility is associated with an inefficient antioxidant defense in rice panicles. Plant Physiol. 2004;121:462–471. http://dx.doi.org/10.1111/j.1399-3054.2004.00341.x

15. Cruz de Carvalho MH. Drought stress and reactive oxygen species. Plant Signal Behav. 2008;3:156–165. http://dx.doi.org/10.4161/psb.3.3.5536

16. Karam F, Masaad R, Sfeir T, Mouzer O, Rouphael Y. Evapotranspiration and seed yield of field grown soybean under deficit irrigation conditions. Agric Water Manage. 2005;75:226–244. http://dx.doi.org/10.1016/j.agwat.2004.12.015

17. Dogan E, Kirnak H, Copur O. Deficit irrigations during soybean reproductive stages and CROPGRO-soybean simulations under semi-arid climatic conditions. Field Crop Res. 2007;103:154–159. http://dx.doi.org/10.1016/j.fcr.2007.05.009

18. Pukacka S, Ratajczak E. Production and scavenging of reactive oxygen species in *Fagus sylvatica* seeds during storage at varied temperature and humidity. J Plant Physiol. 2005;162:873–885. http://dx.doi.org/10.1016/j.jplph.2004.10.012

19. Mishra NP, Mishra RK, Singhal GS. Changes in the activities of antioxidant enzymes during exposure of intact wheat leaves to strong visible light at different temperatures in the presence of protein synthesis inhibitors. Plant Physiol. 1993;102:903–910.

20. Zieslin N, Ben-Zaken R. Peroxidase, phenylalanine ammonia-lyase and lignification in peduncles of rose flowers. Plant Physiol Biochem. 1991;29:147–151.

21. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976;72:248–254. http://dx.doi.org/10.1016/0003-2697(76)90527-3

22. Westgate ME, Peterson CM. Flower and pod development in water-deficient soybean (*Glycine max* L. Merr.). J Exp Bot. 1993;258:109–117. http://dx.doi.org/10.1093/jxb/44.1.109

23. Smirnoff W. The role of active oxygen in the response of plants to water deficit and desiccation. New Phytol. 1993;125:27–58. http://dx.doi.org/10.1111/j.1469-8137.1993.tb03863.x

24. Mittler R, Zilinkas BA. Regulation of pea cytosolic ascorbate peroxidase and other antioxidant enzymes during the progression of drought stress and following recovery from drought. Plant J. 1994;5:397–405. http://dx.doi.org/10.1111/j.1365-313X.1994.00397.x

25. Nazarli H, Zardashti MR, Darvishzadeh R, Mohammadi M. Change in activity of antioxidative enzymes in young leaves of sunflower (*Helianthus annus* L.) by application of super absorbent synthetic polymers under drought stress condition. Aust J Crop Sci. 2011;5:1334–1338.

26. Kausar R, Hossain Z, Makino T, Komatsu S. Characterization of ascorbate peroxidase in soybean under flooding and drought stresses. Mol Biol Rep. 2012;39(12):10573–10579. http://dx.doi.org/10.1007/s11033-012-1945-9

27. Porcel R, Barea JM, Ruiz-Lozano JM. Antioxidant activities in mycorrhizal soybean plants under drought stress and their possible relationship to the process of nodule senescence. New Phytol. 2003;157:135–143. http://dx.doi.org/10.1046/j.1469-8137.2003.00658.x

28. Yoshimura K, Yabuta Y, Ishikawa T, Shigeoka S. Expression of spinach ascorbate peroxidase isozymes in response to oxidative stress. Plant Physiol. 2000;123(1):223–234. http://dx.doi.org/10.1104/pp.123.1.223

29. Ünyayar S, Keles Y, Cekic FÖ. The antioxidative response of two tomato species with different drought tolerances as a result of drought and cadmium stress combinations. Plant Soil Environ. 2005;51:57–64.

30. Hegedüs A, Erdei S, Horvath G. Comparative studies of H_2O_2 detoxifying enzymes in green and greening barley seedlings under cadmium stress. Plant Sci. 2001;160:1085–1093. http://dx.doi.org/10.1016/S0168-9452(01)00330-2

31. Goud PB, Kachole MS. Antioxidant enzyme changes in neem, pigeon pea and mulberry leaves in two stages of maturity. Plant Signal Behav. 2012;7:1258–1262. http://dx.doi.org/10.4161/psb.21584

32. Bai L-P, Sui F-G, Ge T-D, Sun Z-H, Lu Y-Y, Zhou G-S. Effect of drought stress on leaf water status, membrane permeability and enzymatic antioxidant system of maize. Pedosphere. 2006;16:326–332. http://dx.doi.org/10.1016/S1002-0160(06)60059-3

33. Zarei S, Ehsanpour AA, Abbaspour J. The role of overexpression of *P5CS* gene on proline, ascorbate peroxidase activity and lipid peroxidation of transgenic tobacco (*Nicotiana tabbacum* L.) plant under *in vitro* drought stress. J Cell Mol Res. 2012;4:43–49.

34. Devi R, Kaur N, Gupta AK. Potential of antioxidant enzymes in depicting drought tolerance of wheat (*Triticum aestivum* L.). Indian J Biochem Biophys. 2012;49:257–265.

35. Sheoran S, Thakur V, Narwal S, Turan R, Mamrutha HM, Singh V, et al. Differential activity and expression profile of antioxidant enzymes and physiological changes in wheat (*Triticum aestivum* L.) under drought stress. Appl Biochem Biotech. 2015;177:1282–1298. http://dx.doi.org/10.1007/s12010-015-1813-x

Drought, climate change and sustainability of water in agriculture: A roadmap towards the NWRS2

AUTHORS:
Daniel El Chami[1*]
Maroun El Moujabber[2]

AFFILIATIONS:
[1]Department of Agricultural Economics, University of the Free State, Bloemfontein, South Africa
[2]International Centre for Advanced Mediterranean Agronomic Studies, Valenzano, Italy
*Current address: Timac Agro Italia, Ripalta Arpina, Italy

CORRESPONDENCE TO:
Maroun El Moujabber

EMAIL:
elmoujabber@iamb.it

DATES:
Received: 05 Dec. 2015
Revised: 16 Feb. 2016
Accepted: 25 May 2016

KEYWORDS:
water management; irrigation; National Water Resource Strategy; arid and semi-arid areas; South Africa

HOW TO CITE:
El Chami D, El Moujabber M. Drought, climate change and sustainability of water in agriculture: A roadmap towards the NWRS2. S Afr J Sci. 2016;112(9/10), Art. #2015-0457, 4 pages. http://dx.doi.org/10.17159/sajs.2016/20150457

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

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

The frequency and intensity of drought, extreme events and high wind velocities in South Africa are expected to increase in the next century as a result of the changing climate. The National Water Resource Strategy 2 (NWRS2) has set out the general and strategic directions for water resources management in the country for the next 20 years. However, the strategy does not draw a framework tailored specifically for agricultural use, with specific measures and goals. Therefore, to reach sustainability of water in agriculture, four major strategic goals are suggested, on which research institutions can focus and promote through good governance. The strategy emphasises: (1) crop research to find new drought- and heat- tolerant and resistant breeds and varieties; (2) intensified research in agricultural practices; (3) increasing the efficiency of water use within agriculture; and (4) integrating all these strategic goals within a sustainable research framework. Finally, the research calls for rapid action and implementation.

Significance:

- The framework is proposed for stakeholders and policymakers in higher education, agriculture and resources management in South Africa for new research horizons at national level to improve overall agricultural sustainability by 2030 as stipulated by the Millennium Development Goals.

Introduction

South Africa has a huge territory with a diverse range of climates, but is dominated by a semi-arid climate. The historical annual rainfall varies from less than 100 mm per annum in the west to over 1500 mm in the east; the average rainfall is 450 mm per annum. The year 2015 was characterised by unprecedentedly dry and hot weather – the driest it has been in over a century[1] – with maximum temperature and minimum rainfall records measured in different parts of the country[2,3]. The frequency and intensity of drought and heat events are expected to increase over the next century as a result of the changing climate.[4] According to Solomon et al.[5], the effects will be mainly observed on annual temperatures (with a possible increase of 3.4 °C and up to 3.7 °C in spring) and on rainfall (with a decrease of 23% in winter and 13% in spring).

The current extreme conditions in South Africa have had the same impacts described in the literature (e.g. Parry et al.[6]) – such as severe depletion of water resources (ground and surface water) and, consequently, deterioration of water quality. The impacts have been hard on the agricultural sector, which is extremely sensitive to heat and uses over 60% of the total water resources of the country.[7] Consequently, the whole socio-economic structure of the agriculture-based economy is affected, not to mention the impacts on the environmental health of national ecosystems, the threat to national food security because of the reduction of the harvests and the increase in poverty levels. The industrial and urban sectors have also been strongly affected by this drought, given the water restrictions that have been imposed throughout the country.

The National Water Resource Strategy 2 (NWRS2) – published in 2012 by the Department of Water Affairs – has set out the general and strategic directions for water resources management in the country for the next 20 years to ensure equitable growth and development. The NWRS2 has addressed the management framework at national level according to the Integrated Water Resources Management approach.[8] Integrated Water Resources Management is a cross-sectoral and participatory approach for sustainable development, allocation and monitoring of water resources[9], which, according to the Global Water Partnership[10] is central to achieving the Millennium Development Goals. Further, Integrated Water Resources Management will be instrumental for the 2030 Agenda to target the Sustainable Development Goals on water and water-related targets under other goals.[11,12]

Further, the NWRS2 focused on the importance of reliable information for the implementation success and the partnership between research institutions, industries and local communities. However, these strategic guidelines did not draw any roadmap to reach these promising objectives and achieve the sustainability of the local water resources in agriculture. From another side, the scientific community has called to intensify research on drought mitigation and to find ways involving integrated strategies and measures[13] for farming systems to cope with its effects, especially in the arid and semi-arid areas of southern Africa[14].

In agriculture, there are no versatile solutions for sustainable management of natural resources under drought conditions, but the integration of different measures within one strategy would be an effective one. Herein a framework is presented on the role of universities in collaboration with other national research institutions to gather and analyse reliable data and to promote innovation towards the NWRS2 goals through good governance. The suggested strategy is divided into four main dimensions, covering all aspects of agriculture which play a crucial part in the overall sustainability of water in agriculture: crop research, agricultural practices, irrigation management and sustainability research.

Crop research

Combining plant physiology and genetics is a powerful scientific tool that can be used to increase efficiency and productivity in agriculture. Crop research should focus on plant physiology and plant genomics, as well as breeding for drought- and heat- resistant and tolerant traits.

Agriculture in South Africa is a major source of income and a significant portion of the population depends on agricultural activity. Drought and heat conditions are the main constraints in agricultural areas. According to recent research,[15,16] the problem is very likely to become worse if trends in climate change play out as expected. The foremost challenge is to select cost-effective crops and varieties with drought- and heat-tolerant and resistant traits for different South African agro-climate zones.

While water is crucial for plant growth, some crops can cope with drought conditions (drought-tolerant crops), others have the ability to tolerate substantial dehydration of their tissues and organs as well as overheating conditions (drought-resistant crops), while others are resilient to high temperatures (heat-resistant crops). Although the effect of heat stress is influenced by additional stresses such as drought, it is important to acknowledge the individual consequences of each.[17]

There are numerous crops (staples, field crops, legumes, vegetables and trees) that are naturally drought- and heat- tolerant and resistant; however, integrated research in advanced plant physiology, molecular genomics and biology, and breeding can, under similar conditions, create traits that allow crops to grow successfully in drought-prone environments.[18] The integration of new tools and technologies has great potential for improving outcomes in the future[19] when creating new performant traits or finding desirable traits in wild crop relatives adapted to specific conditions. Even though many scientists believe that genetic engineering is not yet an effective tool to provide drought- and heat-tolerant and resistant crops[20,21], this area of research has substantial future potential. In the meantime, crop breeding (especially the conventional breeding using marker-assisted selection) is an extremely useful technique for breeding complex traits (such as drought tolerance) into improved varieties.[22]

There is a dearth of studies reported in the literature to enable quantification of water consumption and savings as a result of such improved varieties. However, a study has shown that drought-tolerant and drought-resistant traits could improve the yield of maize in Africa by up to 30% under drought conditions when compared with commercial seeds.[23] Furthermore, Gur and Zamir[24] successfully obtained 50% higher yields of tomato, even under drought conditions, by breeding cultivated tomatoes with wild relatives. Deryng et al.[25] found – on a global scale – that by the 2080s, extreme heat stress at anthesis could: (1) double losses of maize yield, (2) reduce projected gains in spring wheat yield by half, and (3) reduce projected gains in soybean yield by a quarter.

Agricultural practices

Biodiversity and healthy soil are central to ecological approaches for more drought-resistant and drought-tolerant farming practices, which is more resilient to extreme events.[26] Practices encouraging biodiversity and soil health would increase the available water for crops and improve sustainability of agricultural production.[27]

Many of these agricultural practices are available and are used in different places in the world to manage crops and soils. They aid in building healthy and resilient soils to cope with drought.[28] In the windy, semi-arid climate of South Africa, in which soils are prone to degradation by salinity and erosion, these practices could: (1) improve soil moisture holding capacity, (2) reduce soil erosion and salinity, and (3) increase biodiversity of the system. For instance, polyculture and the use of cover crops and crop residues are efficient practices used to protect soils from erosion. Methods such as conservation tillage, legume intercrops, manure and composts would help build soils rich in organic matter, enhancing soil structure and mulching. These practices enhance water infiltration and retention, and make nutrients more accessible to the plant.[29,30]

Some of these practices have been studied in Pennsylvania (USA)[31] and the results show that, under severe drought conditions, 'organic animal-based' and 'legume-based cropping' produced 28–34% higher yields for corn and 36–50% higher yields for soybean compared with those after conventional cropping.[32] In Colorado (USA), Peterson and Westfall[27] estimated that the soil 'water storage efficiency' varied between 19% for maximum tillage and 40% for minimum tillage. Consequently, the precipitation use efficiency was estimated to be 1.22 kg/ha/mm for maximum tilled winter wheat compared with 3.25 kg/ha/mm for minimum tillage – i.e. a higher yield under minimum tillage.[27]

Irrigation management

Efficient management of irrigation water requires a systematic approach that considers the optimisation of rainfall use, the treatment and reuse of non-conventional sources of water, the efficiency of irrigation systems and the reduction of on-farm irrigation losses. According to the Department of Water Affairs[7], South Africa has almost reached its limit in developing surface-water sources, and, for various reasons, not many more dams can be built to increase the supply. What can be done, however, is to keep the systems (storage and conveyance) adequately maintained so that the conveyance efficiency is increased to maximum levels. Water lost during transport through canals could cause the conveyance efficiency to decrease to 95%.[32] In South Africa, it was estimated that this efficiency for different irrigation schemes varies between 63% and 86%.[33]

Nevertheless, to increase on-farm water availability, reservoirs for rainwater harvesting (where possible) could be an economic opportunity for farmers in remote areas. Reservoir systems could also provide farmers – especially smallholder farmers[36] – with additional water during drought periods[34,35]. Water treatment and water reuse in agriculture – despite the quality concerns that it may raise[37] – could be another source of water for irrigation if managed properly and monitored periodically[38].

Furthermore, not all water taken from a source reaches the field, and the plants do not use all the water arriving at the field. Besides the conveyance efficiency, there is a 'field application efficiency' (EA) which reflects the volume of water effectively used by the plants. The EA depends on the irrigation system used and the level of on-farm management[32]. In theory, EA could vary from 60% for surface irrigation, 70% for sprinkler systems and 90% for drip systems[32], but in South Africa, Reinders[39] estimated that the EA varies between 76% and 82% for different sprinkler and drip systems on different sugarcane fields. Finally, according to Wallace[40], about 44% of the total water resource at source is lost between storage and conveyance, and as run-off and/or drainage in irrigation.

On-farm irrigation management and practices play a crucial part in improving EA and reducing irrigation losses; but several factors need to be considered, mainly the quantity, the quality and the spatial and temporal distribution of water resources.[30] To begin with, water quality, as mentioned earlier, should be subject to strict management and regular monitoring. Furthermore, water quantity encompasses the selection of water efficient irrigation systems that could improve efficiency to 90%.[32] For instance, sprinkler systems with low pressure increase irrigation efficiency substantially under windy conditions.[41] In Montana (USA), Bauder[42] estimated that low-pressure sprinkler systems reduce average water losses by about 50% compared with high-pressure systems. We should not exclude precision agriculture as a practice, as it could reduce water use and improve farm economics considerably, and thereby subsequently contribute to the sustainability of agricultural production.[29,43] Indeed, computerised and GPS-based technologies enable farmers to control more precisely the spatial and temporal management of water application as provided by wireless sensor networks.

The implementation of well-managed deficit irrigation scheduling could contribute considerably to water savings and further improve farmers' benefits without compromising the yields.[44] Further, with scientific advances in physiology, new avenues in the improvement of deficit irrigation could be explored in the future.[45] In addition, irrigation scheduling requires not only the optimisation of water volume, but also the optimisation of irrigation timing. The interaction among wind velocity, temperature and humidity,

changes the EA extensively. According to Bauder[42], EA might vary in sprinkler irrigation between 9% and 26%, depending on the wind velocity, the pressure of the system and the volume of water applied. Moreover, as seen in the semi-arid areas of Spain, by shifting irrigation from day to night, water loss from wind drift and evaporation losses in sprinkler solid-sets and moving laterals could be reduced by up to 50%.[46,47] This shift should be seriously considered in South Africa where it is common practice to irrigate with high-pressure sprinkler systems in the middle of windy days when wind drift and evaporation losses are extremely high.

Sustainability research

Sustainability falls within transformative research. The major concerns of this type of research are to solve fundamental problems in order to secure effective, equitable and durable solutions to agriculture and food production in a changing climate. These difficulties are increasing the pressure on our ecosystems and our societies.

To reach sustainable management of water resources in agriculture, an integrated approach should be applied, which combines the latest findings in crop research, agricultural practices and irrigation management. In this way, the most cost-efficient options that would alleviate environmental burdens can be chosen. El Chami and Daccache[48] have introduced a novel approach to implement sustainability assessment. It consists of the integration of crop growth models, life-cycle assessment models, general circulation models and economic models into one framework for application at farm level. The application of this approach could for example be to compare different drought-resistant varieties of the same crop, under different conservation practices and different irrigation schedules and techniques. This application would increase the options for farmers to optimise resource use while keeping income in mind.

Sustainability research could also focus on integrated modelling for large-scale regional optimisation, taking into account the regional water balance, climate change and land use.[49] Within this approach, the available natural resources (soil, water and air) could be optimised on a life-cycle basis. This modelling could also integrate water treatment and re-use within the regional water balance.

Conclusions

The NWRS2 has been briefly reviewed and a practical framework to follow for the sustainability of the agricultural water resources in South Africa has been presented. This framework would reduce water losses in agriculture, increase water use efficiency, and optimise the use of other natural resources (soil and air), which would benefit farmers. Governmental bodies play an essential role in water management in terms of designing, implementing and monitoring policies that promote sustainable management of national water resources for agricultural use (e.g. charges and taxes, incentives and penalties, as well as banking facilities for new and efficient systems). However, the focus here was on research in agriculture which should be adopted by universities in South Africa in collaboration with other research institutions, especially public policies that encourage funding research in sustainable and transformative science; such research should include[50,51]:

1. Strategic crop research, with a focus on resources and technologies to find new breeds and varieties that perform better under the local drought and heat conditions in order to improve water use efficiency without compromising yields.
2. Intensified research in agricultural practices, including ecological and conservational approaches to assess, for different local crops, the yields and water use efficiency of national agriculture.
3. Research on methods to increase the efficiency of water use within agriculture as an essential and strategic priority.
4. Integration of all of the above within a sustainable research framework at farm and large scales.

The dimensions of this roadmap are designed on a national level. However, for successful governance, regional stakeholders guided by these pillars are invited to design regional steps, which cover the specific needs of local communities. In parallel, the involvement of society should be increased by a participatory approach for an efficient governance. Societal involvement should be accompanied by media awareness campaigns to promote good practices and discourage conventional practices (e.g. soil burning and ploughing). Educational programmes at schools and universities, as well as educational messages (similar to promotional messages) on radio and television, could have high social impact. Extension services should also address scarcity issues explicitly through farmers' magazines, short courses and pilot projects, because recent research shows that farmers have weak incentives to increase farm efficiency through implementing new practices.[52]

Finally, this article is an urgent call for all stakeholders and policymakers in government bodies and in research institutions to accelerate the implementation of the roadmap. It might take years to obtain results, especially because agricultural practices take 2–3 years to establish, and crop research requires, in some cases, decades.

Acknowledgement

The editorial assistance and support provided by Mrs Liesl van der Westhuizen are greatly appreciated.

Authors' contributions

D.E.C. wrote the first draft of the manuscript; M.E.M. made the revisions to the methodology section; both authors approved the final submission.

References

1. Stoddard ED. South African drought follows third-driest season in 80 years. Reuters [edition on the Internet]. 2015 November 10; Environment [cited 2015 Nov 02]. Available from: http://www.reuters.com/article/us-safrica-drought-idUSKCN0SZ1SQ20151110
2. South African Weather Service. General information: Maximum temperature record at Vredendal on 27 October 2015 [document on the Internet]. c2015 [cited 2015 Nov 04]. Available from: http://www.weathersa.co.za/images/documents/280/CLS-CI-GEN-INFO-099.2%20Vredendal%20Max%20Temp%20Record%2027%20Oct%202015.pdf
3. South African Weather Service. Heat wave conditions over Gauteng, Mpumalanga and western Limpopo [media release on the Internet]. c2015 [cited 2015 Nov 05]. Available from: http://www.weathersa.co.za/news-events/media-release/275-heat-wave-conditions-over-gauteng-mpumalanga-and-westerrn-limpopo
4. Rouault M, Richard Y. Intensity and spatial extension of drought in South Africa at different time scales. Water SA. 2003;29(4):489–500. http://dx.doi.org/10.4314/wsa.v29i4.5057
5. Solomon S, Qin D, Manning M, Marquis M, Averyt K, Tignor MB, et al. Climate change 2007 – The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). Cambridge, UK: Cambridge University Press; 2007.
6. Parry ML, Canziani OF, Palutikof JP, Van der Linden PJ, Hanson CE. Climate Change 2007: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). Cambridge, UK: Cambridge University Press; 2007.
7. South African Department of Water Affairs (DWA). Strategic overview of the water sector in South Africa. Pretoria: DWA; 2013.
8. South African Department of Water Affairs (DWA). Proposed National Water Resource Strategy 2: Summary. Pretoria: DWA; 2012.
9. Capacity Building Network for Integrated Water Resources Management (CapNet). Tutorial on basic principles of integrated water resources management [document on the Internet]. c2006 [cited 2015 Nov 11]. Available from: https://www.viawater.nl/tutorial-on-basic-principles-of-integrated-water-resources-management
10. Technical Committee, Global Water Partnership (GWP). How IWRM will contribute to achieving the MDGs. Policy Brief 4. Stockholm: GWP; 2006. Available from: http://www.gwp.org/Global/ToolBox/Publications/Policy%20Briefs/04%20How%20IWRM%20will%20contribute%20to%20achieving%20the%20MDGs%20(2006)%20English.pdf
11. Institute for Global Environmental Strategies (IGES). Placing water at the core of the Sustainable Development Goals (SDGs): Why an integrated perspective is needed. Policy Brief 31. Hayama, Japan: IGES; 2015.

12. Task Force on IWRM, World Water Council (WWC). Integrated water resource management: A new way forward. Marseille: WWC; 2015.
13. Rockström J. Resilience building and water demand management for drought mitigation. Phys Chem Earth. 2003;28(20–27):869–877. http://dx.doi.org/10.1016/j.pce.2003.08.009
14. Seymour C, Desmet P. Coping with drought – Do science and policy agree? S Afr J Sci. 2009;105(1/2):18–19.
15. Hoffman MT, Carrick PJ, Gillson L, West AG. Drought, climate change and vegetation response in the succulent Karoo, South Africa. S Afr J Sci. 2009;105:54–60.
16. Vetter S. Drought, change and resilience in South Africa's arid and semi-arid rangelands. S Afr J Sci. 2009;105:29–33.
17. Bita CE, Gerats T. Plant tolerance to high temperature in a changing environment: Scientific fundamentals and production of heat stress-tolerant crops. Front Plant Sci. 2013;4:273. http://dx.doi.org/10.3389/fpls.2013.00273
18. Cattivelli L, Rizza F, Badeck F-W, Mazzucotelli E, Mastrangelo AM, Francia E, et al. Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. Field Crop Res. 2008;105:1–14. http://dx.doi.org/10.1016/j.fcr.2007.07.004
19. Ashraf M. Inducing drought tolerance in plants: Recent advances. Biotechnol Adv. 2010;28(1):169–183. http://dx.doi.org/10.1016/j.biotechadv.2009.11.005
20. Finkel E. Making every drop count in the buildup to a blue revolution. Science. 2009;323(5917):1004–1005. http://dx.doi.org/10.1126/science.323.5917.1004
21. Collins NC, Tardieu F, Tuberosa R. Quantitative trait loci and crop performance under abiotic stress: Where do we stand? Plant Physiol. 2008;147(2):469–486. http://dx.doi.org/10.1104/pp.108.118117
22. Anami S, De Block M, Machuka J, Van Lijsebettens M. Molecular improvement of tropical maize for drought stress tolerance in sub-Saharan Africa. Crit Rev Plant Sci. 2009;28(1–2):16–35. http://dx.doi.org/10.1080/07352680802665305
23. Gilbert N. Cross-bred crops get fit faster: Genetic engineering lags behind conventional breeding in efforts to create drought resistant maize. Nature. 2014;513:292. http://dx.doi.org/10.1038/513292a
24. Gur A, Zamir D. Unused natural variation can lift yield barriers in plant breeding. PLoS Biol. 2004;2(10), e245, 6 pages. http://dx.doi.org/10.1371/journal.pbio.0020245
25. Deryng D, Conway D, Ramankutty N, Price J, Warren R. Global crop yield response to extreme heat stress under multiple climate change futures. Environ Res Lett. 2014;9, Art. #034011, 13 pages. http://dx.doi.org/10.1088/1748-9326/9/3/034011
26. Tirado R, Cotter J. Ecological farming: Drought-resistant agriculture. GRL-TN 02/2010. Amsterdam: Greenpeace International; 2010. Available from: http://www.greenpeace.org/international/Global/international/publications/agriculture/2010/Drought_Resistant_Agriculture.pdf
27. Peterson GA, Westfall DG. Managing precipitation use in sustainable dryland agroecosystems. Ann Appl Biol. 2004;144(2):127–138. http://dx.doi.org/10.1111/j.1744-7348.2004.tb00326.x
28. Habiba U, Shaw R, Takeuchi Y. Farmer's perception and adaptation practices to cope with drought: Perspectives from Northwestern Bangladesh. Int J Disaster Risk Reduction. 2012;1:72–84. http://dx.doi.org/10.1016/j.ijdrr.2012.05.004
29. Evans RG, Sadler EJ. Methods and technologies to improve efficiency of water use. Water Resour Res. 2008;44(7), W00E04, 15 pages. http://dx.doi.org/10.1029/2007WR006200
30. Deng X-P, Shan L, Zhang H, Turner NC. Improving agricultural water use efficiency in arid and semiarid areas of China. Agr Water Manage. 2006;80(1–3):23–40. http://dx.doi.org/10.1016/j.agwat.2005.07.021
31. Pimentel D, Hepperly P, Hanson J, Douds D, Seidel R. Environmental, energetic, and economic comparisons of organic and conventional farming systems. BioScience. 2005;55(7):573–582. http://dx.doi.org/10.1641/0006-3568(2005)055[0573:EEAECO]2.0.CO;2
32. Brouwer C, Prins K, Heibloem M. Irrigation water management: Irrigation scheduling. Training Manual No. 4. Rome: Food and Agriculture Organization of the United Nations (FAO); 1989.
33. Reid PCM, Davidson DCR, Kotze T. A note on practical methods of improving the conveyance efficiency on a government irrigation scheme. Water SA. 1986;12(2):89–91.
34. Oweis T, Pala M, Ryan J. Stabilizing rainfed wheat yields with supplemental irrigation and nitrogen in a Mediterranean climate. Agron J. 1998;90(5):672–681. http://dx.doi.org/10.2134/agronj1998.00021962009000050017x
35. Oweis TY, Prinz DP, Hachum AY. Rainwater harvesting for agriculture in the dry areas. London: CRC Press; 2012. http://dx.doi.org/10.1201/b12351
36. Rockström J, Barron J, Fox P. Rainwater management for increased productivity among small-holder farmers in drought prone environments. Phys Chem Earth. 2002;27(11–12):949–959. http://dx.doi.org/10.1016/S1474-7065(02)00098-0
37. Toze S. Reuse of effluent water – benefits and risks. Agr Water Manage. 2006;80(1–3):147–159. http://dx.doi.org/10.1016/j.agwat.2005.07.010
38. Rusan MJM, Hinnawi S, Rousan L. Long term effect of wastewater irrigation of forage crops on soil and plant quality parameters. Desalination. 2007;215(1–3):143–152. http://dx.doi.org/10.1016/j.desal.2006.10.032
39. Reinders FB. Performance of irrigation systems and the impact on water use efficiency. Paper presented at: SASIAA Annual Symposium 2001: Water Use Efficiency: Potential and Constraints. Available from: http://www.sasa.org.za/Libraries/SA_Sugarcane_Industry_Agronomists_Association/Performance_of_irrigation_systems_and_the_impact_on_water_use_efficiency_Felix_Reinders.sflb.ashx
40. Wallace JS. Increasing agricultural water use efficiency to meet future food production. Agr Ecosyst Environ. 2000;82(1–3):105–119. http://dx.doi.org/10.1016/S0167-8809(00)00220-6
41. Dukes MD. Effect of wind speed and pressure on linear move irrigation system uniformity. Appl Eng Agric. 2006;22(4):541–548. http://dx.doi.org/10.13031/2013.21222
42. Bauder J. How much irrigation water do you lose when it's windy [homepage on the Internet]. c2000 [cited 2015 Nov 27]. Available from: http://www.montana.edu/cpa/news/wwwpb-archives/ag/baudr258.html
43. Bongiovanni R, Lowenberg-Deboer J. Precision agriculture and sustainability. Precis Agric. 2004;5(4):359–387. http://dx.doi.org/10.1023/B:PRAG.0000040806.39604.aa
44. Fereres E, Soriano MA. Deficit irrigation for reducing agricultural water use. J Exp Bot. 2007;58(2):147–159. http://dx.doi.org/10.1093/jxb/erl165
45. Sinclair TR, Purcell LC. Is a physiological perspective relevant in a 'genocentric' age? J Exp Bot. 2005;56(421):2777–2782. http://dx.doi.org/10.1093/jxb/eri297
46. Yacoubi S, Zayani K, Zapata N, Zairi A, Slatni A, Salvador R, et al. Day and night time sprinkler irrigated tomato: Irrigation performance and crop yield. Biosyst Eng. 2010;107(1):25–35. http://dx.doi.org/10.1016/j.biosystemseng.2010.06.009
47. Playán A, Salvador R, Faci JM, Zapata N, Martínez-Cob A, Sánchez I. Day and night wind drift and evaporation losses in sprinkler solid-sets and moving laterals. Agr Water Manage. 2005;76(3):139–159. http://dx.doi.org/10.1016/j.agwat.2005.01.015
48. El Chami D, Daccache A. Assessing sustainability of winter wheat production under climate change scenarios in a humid climate – An integrated modelling framework. Agr Syst. 2015;140:19–25. http://dx.doi.org/10.1016/j.agsy.2015.08.008
49. D'Agostino DR, Scardigno A, Lamaddalena N, El Chami D. Sensitivity analysis of coupled hydro-economic models: Quantifying climate change uncertainty for decision-making. Water Resour Manag. 2014;28(12):4303–4318. http://dx.doi.org/10.1007/s11269-014-0748-2
50. South African National Research Foundation (NRF). Strategic plan 2015–2020. Pretoria: NRF; 2015.
51. South African Water Research Commission (WRC). Key strategic areas (KSA's). Pretoria: WRC; 2015. Available from: http://www.wrc.org.za/Pages/Research.aspx
52. Levidow L, Zaccaria D, Maia R, Vivas E, Todorovic M, Scardigno A. Improving water-efficient irrigation: Prospects and difficulties of innovative practices. Agr Water Manage. 2014;146:84–94. http://dx.doi.org/10.1016/j.agwat.2014.07.012

An extended Farm Site Development Method

AUTHORS:
Eugene Hanekom[1]
Andre Liebenberg[1]
Marné de Vries[1]

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

CORRESPONDENCE TO:
Marné de Vries

EMAIL:
marne.devries@up.ac.za

DATES:
Received: 12 Nov. 2015
Revised: 17 Apr. 2016
Accepted: 03 June 2016

KEYWORDS:
method engineering; farm facilities planning; financial planning; optimisation; strategic planning

HOW TO CITE:
Hanekom E, Liebenberg A, De Vries M. An extended Farm Site Development Method. S Afr J Sci. 2016;112(9/10), Art. #2015-0427, 11 pages. http://dx.doi.org/10.17159/sajs.2016/20150427

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
University of Pretoria

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

The inefficient and ineffective use of arable land in South Africa is one of the numerous challenges within its agricultural sector. Previous research has indicated that a method, the Farm Site Development Method (FSDM), could increase the effective and efficient use of arable land by providing a roadmap to the farm owner for incrementally transforming the current state facilities and resources of a farm towards a future saturation state. The FSDM was then demonstrated at a crop-producing farm and several opportunities existed to extend its utility. Here we suggest its extension for application to a livestock farm, and also include optimisation techniques, demand planning and financial planning.

Significance:

- Extension of the Farm Site Development Method, and its demonstration at a livestock farm, for facilities and financial planning within the agricultural sector.

Introduction and background

The primary agricultural sector in South Africa is growing.[1] Yet there has been a drop in GDP contribution from 7.1% in 1970 to 2.6% in 2013.[2] Numerous factors influence the slow growth of this sector, such as climate change[3], the land reform projects of which at least 50% failed[4], and inefficient use of arable land[5]. Previous work has already identified an opportunity to improve the efficient and effective use of arable land by introducing long-term planning techniques to the agricultural sector. A new artefact was developed, called the Farm Site Development Method (FSDM), as a roadmap for incrementally transforming the current state facilities and resources of an existing farm towards a future saturation state that is financially feasible. The FSDM was evaluated with a demonstration at a crop-producing farm.[6]

Although the FSDM was useful to a crop-growing farm, several deficiencies were identified during the demonstration of the method: (1) optimisation techniques were not used to optimise product mix, (2) demand planning was not incorporated to forecast production requirements, and (3) financial planning did not consider cash flow.[6] In addition, drawing knowledge from the field of method engineering, we postulated that the utility of the FSDM could be further improved. The existing FSDM could also be further extended for use on other types of farms, such as livestock farms.

Background literature

Previous research combined existing techniques, i.e. facilities planning and incremental design, to the agricultural sector. Because current facilities or resources may be depleted when the farming enterprise grows, active planning is required for their replacement or extension. The FSDM was developed to evolve the farm facilities in a phased approach towards its future or saturation state.[6]

The FSDM

The FSDM comprises eight sequential steps[6]:

1. Analyse the current-state facility layout.
2. Calculate the saturation state for the farm (compile a saturation-state facility layout).
3. Determine the production requirements and the saturation date.
4. Identify critical resources, utilities, services and/or structures (RUSS) and the design criteria.
5. Identify and evaluate alternatives for RUSS replacement or extension.
6. Compile a series of phase plans, called the farm site development plan (FDP).
7. Represent phase plans graphically in support of the FDP.
8. Validate the FDP.

Each step is elaborated in more detail in Van der Merwe et al.[6] As the FSDM could be classified as a method artefact, the field of method engineering applies and will be discussed next.

Method engineering, situational method engineering and utility criteria

Method engineering has developed within the information systems discipline as a field concerned with the construction of new methods from existing methods.[7] One of the areas in method engineering – situational method engineering – has the objective of constructing methods which are tuned to specific situations and project types.[8,9] Whereas a new method is constructed according to requirements, *situational method engineering* requires adaptation of a generic method according to project-specific needs.[10] Because this article focuses on the extension of an existing method (the FSDM), guidelines from situational method engineering were applicable in adapting the generic FSDM for different situations, i.e. a crop-producing farm versus a livestock farm.

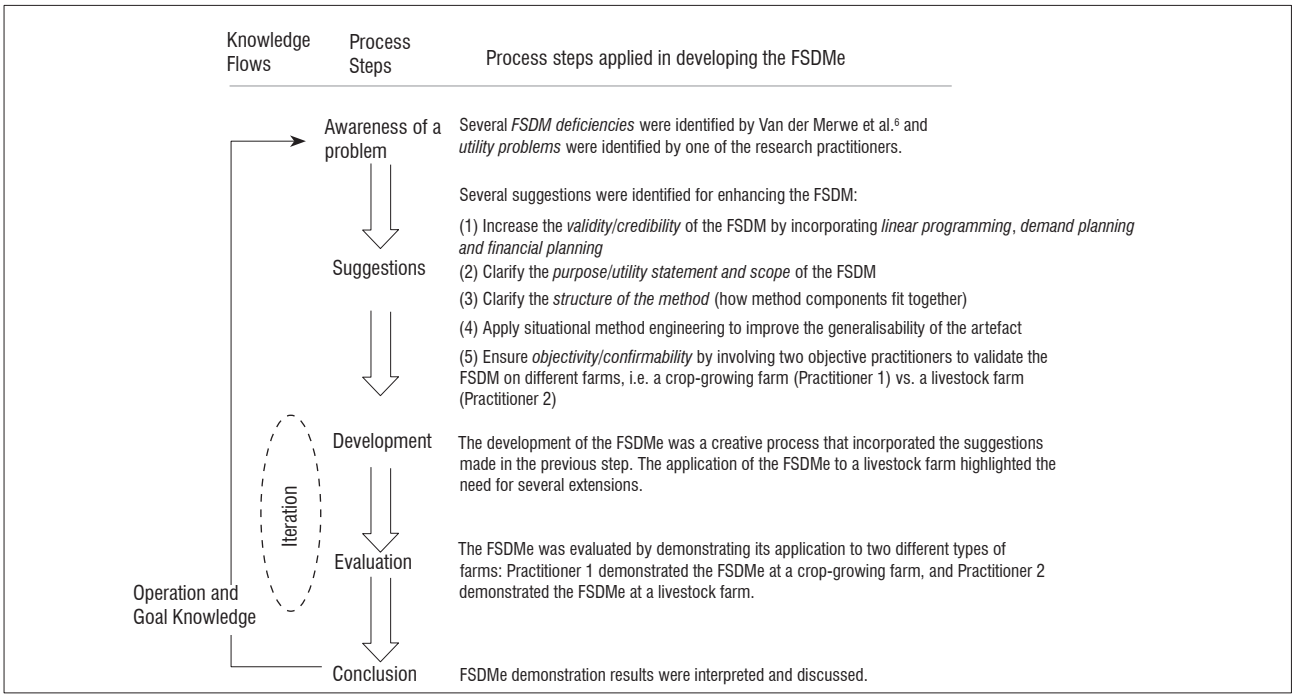


Figure 1: Applying a second design cycle for developing the extended Farm Site Development Method (FSDMe).

Based on the structure specified by Gregor and Jones[11] for design theory, Offerman et al.[10] consider *utility* as the primary evaluation criterion for evaluating a design theory. They define four criteria in terms of utility: (1) *validity*/*credibility* as the ability of a method to yield the claimed utility within the intended scope, (2) *objectivity*/*confirmability* as the degree to which others are able to confirm the utility statement, (3) *generalisability* as the assumption that the utility statement holds true for all possible instances within the scope, and (4) *transferability* as the postulation that the utility statement would also be partially true for out-of-scope situations. From the four criteria, Offerman et al.[10] developed guiding questions for the components of a method theory, which we have applied below to identify utility deficiencies within the existing FSDM. We also evaluate the enhanced FSDM in terms of the four utility criteria of Offerman et al.[10]

First we introduce techniques that could be applied to extend the existing FSDM: linear programming, demand planning and financial planning.

Techniques to enhance the FSDM

Linear programming is a deterministic, optimising model that maximises or minimises a quantifiable objective. The objective could be expressed as a linear function of identified decision variables. The purpose is to optimise the objective by calculating values for the decision variables, given certain constraints. The constraints are also expressed as linear equations or linear inequalities.[12]

Demand planning deals with uncertainties within the business environment when predicting future demand of products. Time series analysis is useful if a relationship exists between the historical demand and future demand of the product, i.e. that past history of demand levels will influence the future behaviour of this demand over time.[12] Historical demand data could be used to identify the average demand for a period, a trend, seasonal elements, cyclical elements, random variation and autocorrelation. Based on the analysis, different time series forecasting techniques may be useful.[13]

Financial planning concerns the optimal use of financial resources and should be performed on two levels – a strategic level and a tactical level. Strategic planning is concerned with long-term goals in the enterprise, such as product mix and new production facilities, whereas tactical planning ensures that existing resources are exploited. The end result of financial planning is to maximise the ratio of profits to the equity capital (i.e. capital used to acquire assets) employed.[12] The financial planner needs to estimate income and costs, differentiating between variable and fixed costs. Seal et al.[14] define variable costs as a cost type that varies in direct production to the level of activity (e.g. fertiliser for crop production), whereas fixed costs remain constant regardless of the level of activity (e.g. property tax).

Research design

The research design used to extend the FSDM – design research – values the creation of novel artefacts and provides a body of knowledge to guide the researcher during different stages of a design cycle.[15] Previous research applied design research to develop a new method artefact, the FSDM. During the first iteration of the design cycle, the validity/credibility of the FSDM was evaluated by applying the FSDM to a crop-producing farm named Waterfall Farm. Several recommendations were made for enhancing the FSDM.[6] Here we apply a second design cycle, using the five-step design cycle of Vaishnavi and Kuechler[16] (see Figure 1, Process Steps), to develop an extended FSDM, called FSDMe.

Referring to Figure 1, 'Awareness of a problem' entailed identification of FSDM deficiencies and utility problems, which led to several 'Suggestions' that were incorporated during a creative 'Development' process for developing the FSDMe. The FSDMe underwent 'Evaluation' by demonstrating its application to a crop-growing farm, as well as a livestock farm. As previous work already demonstrated the FSDM on a crop-growing farm,[6] here we present the demonstration of the FSDMe only on a livestock farm. Although we present the FSDMe as a theoretical contribution in this article, we have also developed a software application, using MS Excel, which presents a user-friendly interface to the user (e.g. practitioner or farm owner) of the FSDMe. The software application is available on request.

The extended Farm Site Development Method: FSDMe

The FSDMe is a method that is used to evolve farm facilities in a phased approach towards its future or saturation state. The FSDMe comprises nine steps and the structure (as depicted in Figure 2) is as follows:

- The purpose (utility-statement), scope, prerequisites, roles and input requirements/data of the FSDMe.
- The method steps and extension points to accommodate different contexts.

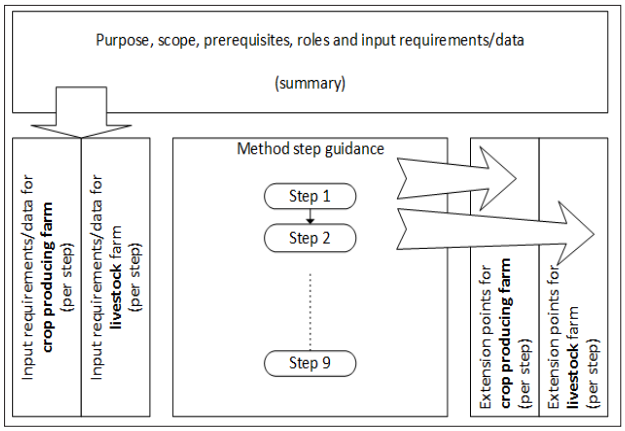


Figure 2: The extended Farm Site Development Method – FSDMe.

Purpose, scope, prerequisites, roles and input requirements/data

Purpose: The FSDMe is a strategic planning tool that facilitates production growth on an existing farm of fixed size and the incremental extension of its facilities towards the saturation state. The main output of this method is a FDP, which accommodates financial and other criteria.

Scope: The FSDMe has been demonstrated on a crop-producing farm as well as a livestock farm. For the crop-producing farm, the FSDMe requires choice of crop products as input, but does not aim to optimise the production mix proportions. For a livestock farm, the FSDMe similarly requires a choice of animal type as input, but also optimises the proportion of animal types. Furthermore, it is assumed that the livestock farm produces crops for the sole purpose of feeding livestock. For a livestock farm we also assume that calves and lambs are only feeding from their mother and do not need to graze. If additional feeding is required, calves and lambs receive creep feed (a blend of maize, molasses, proteins and micro- and macro-minerals). We also assume that calves are sold once they are weaned.

Prerequisites: Decisions about crop varieties and crop production mix should have been made (based on market analyses and soil analyses) prior to using the FSDMe.

Roles: The *farm owner* or the *practitioner* will be applying the FSDMe to generate the FDP. The *farm owner* or *practitioner* should have practical knowledge of the farming industry and the farm produce. As an example, the owner or practitioner of a livestock farm needs to have skills pertaining to breeding techniques, feeding of livestock, knowledge about natural grazing and livestock diseases.

Input requirements: See Table 1.

Method steps with extension points

Step 1: Analyse the current-state facility layout

Input: Existing farm layout plans

Method step guidance: Conduct a site visit to confirm and supplement existing layout plans, photos, GPS coordinates and land surveys. If available, existing plans should be used to draw the current-state facility plan using an appropriate drawing program, such as AutoCAD. If no existing layout plans are available, use GPS coordinates to map the current-state facility layout, indicating all buildings and site features to scale.

Step 2: Calculate the arable land

Input: Total area of the farm (ha)

Method step guidance: Calculate the arable land (a) of the farm (in hectares) as total area less all existing facilities, production areas and road areas.

Table 1: Input requirements/data for using the extended Farm Site Development Method (FSDMe)

Inputs irrespective of production type:	
If market demand is a constraint, historical production data are required. Farmer-prioritised investment decisions (e.g. the decision to buy hail shields). Cash used for expansion (as a percentage of total profits).	
Inputs for crop production	**Inputs for livestock production**
Farm	**Farm**
Existing farm layout plans Total area of the farm (in ha)	Existing farm layout plans Total area of the farm (in ha) Section sizes (ha) ……(a_1, a_2,…a_k)
Produce	**Produce**
Varieties produced and proportion (%) production for each variety Crop growth times for reaching maturity (weeks) Planting density (in units per ha) Waste or cut rate (%)	Choice of animal type *i* Choice of feed crops Equivalent livestock units (in LU*) for each animal type Selected breeding technique (insemination or synchronisation or natural breeding) and number of females (f_i) associated with 1 male during breeding for animal type *i* Production potential (in offspring/year) of each female animal type *i* ……(o_1, o_2,…o_i) Section yield (LU/ha/year) ……(u_1, u_2,…u_k)
Budget	**Budget**
Variable costs (seeds, fertiliser, planting, harvesting, electricity, etc.) and associated inflation rates Fixed costs (asset depreciation, taxes, etc.) and associated inflation rates Product selling price and associated inflation rates	Variable costs (animal feed, veterinary, etc.) and associated inflation rates Fixed costs (e.g. asset depreciation, taxes, etc.) and associated inflation rates Auction price of animal type *i* (when selling) and associated inflation rate Auction price of animal type *i* (when buying) and associated inflation rate Proportion of old animals that are sold for animal type *i*

**LU is a standard unit used to measure the number of animals that will be able to graze on one hectare annually. A fully grown cow of 450 kg = 1 LU*

Step 3: Calculate the yield potential to generate a facility layout at saturation state

Input: Arable land (a) from previous step

Inputs – crop-producing farm:

- Varieties produced and percentage production for each variety
- Crop growth times for reaching maturity (weeks)
- Planting density (units per ha), u
- Waste or cut rate (in percentage), d

Inputs – livestock farm:

- Choice of animal type (note: production mix proportions are not required), i
- Choice of feed crops per farm section
- Equivalent livestock units (in LU) for each animal type
- Selected breeding technique (insemination or synchronisation or natural breeding) and number of females (f_i) associated with one male during breeding for animal type i
- Production potential (in offspring/year) of each female animal type i ……$(o_1, o_2,\ldots,o_i)$
- Section sizes (ha) ……$(a_1, a_2,\ldots,a_k)$
- Section yield (LU/ha/year) ……$(u_1, u_2,\ldots,u_k)$
- Variable costs (animal feed, veterinary, etc.) and associated inflation rates
- Auction price of animal type i (when selling) and associated inflation rate
- Auction price of animal type i (when buying) and associated inflation rate

Method step guidance: The objective is to estimate the yield potential (y) of the farm at saturation state. Based on the yield potential, use a drawing tool to represent the facility layout at the saturation state.

Extension – crop-producing farm:

Yield potential on a crop-producing farm is calculated by:

$$y_c = a * u, \qquad \text{Equation 1}$$

where y_c is the crop yield potential (units), a is the arable land size (ha) and u is the maximum surface utilisation, i.e. maximum number of crop units that could be grown per hectare.

The yield potential at saturation state could also be converted to potential production rate (p_u) in units per week and potential production rate (p_s) in standardised units per week.

$$\text{Therefore: } p_u = \frac{y}{x}(1-d), \qquad \text{Equation 2}$$

where x is the *average* growth timeframe of all crop varieties to reach maturity (in weeks), also incorporating seasonality and growth times for different crop varieties, d is the *average* loss of production units (in units) before distribution and y is the yield potential (units), calculated in Equation 1.

$$\text{Expressed in standardised units: } p_s = \frac{p_u}{s}, \qquad \text{Equation 3}$$

where p_u is the potential production rate in units per week, calculated in Equation 2, and s is the number of production units per standardised unit, where the standardised unit is a packaging unit for transporting the units (produce).

Extension – livestock farm:

Yield potential on a livestock farm is calculated by:

$$y_l = \sum_{j=1}^{k} a_j * u_j, \qquad \text{Equation 4}$$

where y_l is the yield potential (LU per year), a is the section size (ha)…… $(a_1, a_2,\ldots a_k)$ and u is the section yield (LU/ha/year) ……$(u_1, u_2,\ldots u_k)$.

The yield potential could also be expressed as the number of animals of a particular type (L_i). A linear programming model is used to calculate the values for L_i to maximise profits (Z). Because not all animals are delivering offspring, a correction factor ($fi/(fi+1)$) reduces the income to account for the male individuals within the mix.

Therefore: $Max\ Z = \sum_{i=1}^{n}(f_i/(f_i+1))Li(Pi\text{-}Ki)$

$$i \in I = \begin{matrix} \textit{Animal type } 1 \\ \textit{Animal type } 2 \\ \textit{Animal type } n \end{matrix}$$

The constants are:

- y_l = yield potential (LU per year), calculated in Equation 4
- X_i = ratio animals of type i equivalent to 1 LU, where 1 cow = 1 LU, and $i \in I$
- f_i = number of females associated with 1 male according to the breeding technique for animal type i
- P_i = the income generated from female animal type $i \in I$ *in one year* i.e. (selling price for animal type i)*(offspring/year o_i for animal type i)
- K_i = the cost to keep animal type $i \in I$ on the farm *for one year*

The variables are:

- L_i = The number of animals of type i to keep on the farm *annually,* where $i \in I$

Subject to:

- $\sum_{i=1}^{n} \frac{Li}{Xi} \leq y_l$ (saturation state LU may not exceed the LU capacity of the farm)
- $X_iL_i = X_i + L_{i+1}\ \forall i = 1,2,3, \ldots, i\text{-}1$ (relationship of different animal types on the farm remains constant)

As L_i specifies the total number of animals of type i, the breeding technique ratio should be used to calculate $L_{i,f}$ (female animals of type i) and $L_{i,m}$ (male animals of type i):

$$L_{i,f} = (f_i/f_i+1)L_i \qquad \text{Equation 5}$$

$$L_{i,m} = L_i - L_{i,f} \qquad \text{Equation 6}$$

Step 4: Determine the production requirements and the saturation date based on constraints

Inputs:

- If market demand is a constraint, historical production data are required
- Farmer-prioritised investment decisions (e.g. the decision to buy hail shields)
- Cash used for expansion (in percentage of total profits)

Inputs – crop-producing farm:

- Variable costs (e.g. seeds, fertiliser, planting, harvesting, electricity) and associated inflation rates
- Fixed costs (e.g. asset depreciation, taxes) and associated inflation rates
- Product selling price and associated inflation rates

Inputs – livestock farm:

- Variable costs (feeding crops etc.) and associated inflation rates
- Fixed costs (e.g. asset depreciation, taxes, etc.) and associated inflation rates
- Auction price of animal type *i* (when selling) and associated inflation rate
- Auction price of animal type *i* (when buying) and associated inflation rate
- Percentage of old animals that are sold for animal type *i*

Method step guidance: Because the historical resource-acquisition decisions, existing cost structure and other investment decisions limit the cash available for production growth, this method acknowledges the need for incorporating cash flow calculations. The method is simplified by assuming that a fixed percentage of profits is used to expand production on the farm. The farmer also needs to determine if there is a demand constraint for the products. If a demand constraint exists, historical production data should be analysed to project future demand, using time series analysis. The saturation date is calculated as the date on which the saturation state is achieved, which gives an indication of the planning horizon for extending production on the farm. The projected growth of production (per product type) up to the saturation date could be demonstrated graphically.

Step 5: Identify critical RUSS and design criteria

Method step guidance: The specific farming industry should be considered when determining the most important RUSS required for the FDP, as well as the appropriate design criteria (e.g. financial and technical criteria) for evaluating alternatives pertaining to the RUSS. Based on the needs of the farm owner, additional criteria may have to be incorporated from best practice frameworks such as the Global GAP (Good Agricultural Practices), Farming for the Future, the Food Technical Standard and Protocol of the British Retail Consortium (BRC), Bird Friendly standards and Fair Trade standards.

Step 6: Identify and evaluate alternatives for RUSS replacement or extension

Input: Design criteria and RUSS identified in the previous step.

Inputs – crop-producing farm:

- Potential production rate (p_s) in standardised units per week, calculated in Equation 3

Method step guidance: A number of calculations is required to identify alternative restoration initiatives and to evaluate the alternative initiatives against the design criteria identified in Step 5.

- Determine the initial RUSS sizes (see 'Extension for crop-producing farm' below).
- Determine the first capacity depletion date for each of the RUSS. Use the expected production requirements or demand (calculated in Step 3) to determine when the capacity of specific RUSS will be depleted for the first time.
- Use the first capacity depletion date to determine the incremental restoration initiatives for each of the RUSS. Incremental restoration initiatives specify the size and quantity of the additional RUSS, based on the standard sizes available in industry and the increase in demand. Because the restored capacity of a resource or structure may be depleted several times within the planning horizon of the FDP, several incremental restoration initiatives will be required for each of the RUSS.
- Determine the restoration dates for the incremental restoration initiatives. Consider the design criteria (e.g. financial and technical) lead times for constructing and/or acquiring RUSS to complete restoration before capacity depletion occurs.

Extension for crop-producing farm:

Determine the initial RUSS sizes expressed in standardised units. Calculate the RUSS capacity that is currently used, given the existing production rate.

The existing production rate, expressed in standardised units is:

$$p'_s = \frac{p'_u}{s}, \qquad \text{Equation 7}$$

where p'_u is the current production rate in units per week and s is the number of production units per standardised unit, where the standardised unit is a packaging unit for transporting the units (produce).

Step 7: Compile a series of phase plans: The FDP

Method step guidance: Considering the practicality of restoration activities, the practitioner needs to group the restoration dates in a series of phases (from one month to a year) for the entire planning horizon of the FDP. Phase identification enables budgeting and planning for each phase ahead of time.

Step 8: Represent phase plans graphically in support of the FDP

Method step guidance: Draw each phase of the FDP sequentially, starting at the current-state facility layout and ending with a saturation-state facility layout. Although the saturation-state facility layout may resemble the initial facility layout compiled in Step 3, a revised layout may be required to reflect changes in strategy, target market, product or technology.

Step 9: Validate the FDP

Method step guidance: The purpose of this step is to evaluate the sensitivity of the FDP in terms of the quantitative assumptions that were made during the development of the restoration alternatives for the critical RUSS. During this step, the practitioner assesses the rigour of the FDP when different input parameters (e.g. cost of labour) are used or major disasters (e.g. field fires and droughts) are considered. Additional qualitative validation may also be required to ensure that the FDP is useful to management.

Evaluation results: A practical demonstration at Bloemhoek Farm

We present a demonstration of the FSDMe at a livestock farm, which has been documented comprehensively in Hanekom[17].

Input requirements/data

From Table 1 in the previous section, several inputs are required for applying the FSDMe steps. In this demonstration, we provide an interpretation of the inputs that were obtained for generating an FDP for Bloemhoek Farm in Table 2.

Applying the method steps

Applying the method step guidance given in the previous section, we demonstrate its application in this section.

Step 1: Analyse the current-state facility layout

According to Step 1, several inputs are required, which are declared in Table 2.

Method step application: The total area of the farm is divided into different feed yield sections (Figure 3). The yield capacity of each feed yield section is given in livestock unit per hectare (LU/ha) (Table 3), which is a standard unit used to measure the number of animals that would be able to graze in 1 ha annually. A fully grown cow of 450 kg is equivalent to one livestock unit. Table 3 also indicates the current crops planted in each section, as well as the demarcation of Sections 5, 6 and 7 into three grazing camps.

Table 2: Input requirements/data for using the extended Farm Site Development Method (FSDMe) at Bloemhoek Farm

Input parameter	Input value at Bloemhoek Farm
Inputs irrespective of production type	
If market demand is a constraint, historical production data are required	The market demand for meat (beef and mutton) is larger than the potential production capacity of Bloemhoek Farm. Thus, the market demand is not a constraint.
Farmer-prioritised investment decisions	The farmer did not prioritise any investment decisions.
Cash used for expansion (% of total profits)	40% of the gross profit will be allocated to growth of herds.
Farm	
Existing farm layout plans	Existing farm layouts are presented in Hanekom[17] and summarised in Figure 3 and Table 3.
Total area of the farm (in ha)	564.34 ha
Section sizes (ha) (a_1, a_2,…a_k)	Summarised in Table 3.
Produce	
Choice of animal type *i*	The owner decided to farm with type 1 = *cattle* and type 2 = *sheep*. The rationale for this choice is that maximum utilisation of natural grazing could be achieved when rotational grazing is applied, because cattle and sheep consume different lengths of grass.
Choice of feed crops	Lucerne, maize, wheat and weeping lovegrass (*Eragrostis curvula*). The nature of the soil, the nutritive value of the crops, the yield potential and the crop production costs were the main concerns for the selected feed crops.
Equivalent livestock units (in LU) for each animal type	Cattle, weighing 450 kg = 1 LU Sheep, weighing 55 kg = 0.125 LU
Selected breeding technique (insemination or synchronisation or natural breeding) and number of females (f_i) associated with 1 male during breeding for animal type *i*	Cattle – natural breeding, $f_1 = 25$ Sheep – synchronisation, $f_2 = 150$
Production potential (in offspring/year) of each animal type *i* ……(o_1, o_2,…o_i)	For Type 1 (cow): 0.67 For Type 2 (ewe): 1.5
Feed crop yield (LU/ha/year)……(u_1, u_2,…u_k)	Lucerne: 1.9 Maize and wheat: 3.1 Weeping lovegrass (*Eragrostis curvula*): 0.7 Natural grass, shrubs and bush: 0.5 Mountainous terrain: 0.4
Budget	
Variable costs (feeding crops etc.) and associated inflation rates	Annual cost per cow: R2370.74 Annual cost per ewe: R411.97
Fixed costs (e.g. asset depreciation, taxes, etc.) and associated inflation rates	The assumption was made that 40% of gross profit would be used for livestock growth, whereas the remaining 60% would be sufficient to cover fixed costs, such as taxes and asset depreciation. Future costs (feed-processing equipment, livestock-handling facilities and sheep sleeping camps) were also estimated and incorporated during cash flow calculations. Building cost inflation rate: 6.05% (Triami Media, 2014)
Auction price of animal type *i* (when selling) and associated inflation rate	Price per cow (old cow or weaner): R4200 Price per ewe (old ewe or weaner): R615 Inflation rate: 10%
Auction price of animal type *i* (when buying) and associated inflation rate	Price per cow (for breeding): R10 000 Price per ewe: R1500 Inflation rate: 10%
Proportion of old animals that are sold for animal type *i*	% of old cattle sold: 0.01 % of old sheep sold: 0.1

Table 3: Size, current crops and demarcation of grazing camps on Bloemhoek Farm

	Year	2014	2015	2016	2017	2018
Cattle	Number of animals	100	100	100	100	100
	Old animals sold	1	1	1	1	1
	Auction price	R4200.00	R4620.00	R5082.00	R5590.20	R6149.22
	Expenses	R2370.74	R2607.82	R2868.60	R3155.46	R3471.00
	Replacement cost	R10 000.00	R11 000.00	R12 100.00	R13 310.00	R14 641.00
	Total income	R274 433.33	R301 696.31	R331 972.63	R365 633.18	R403 281.71
	Total expenses	R237 074.25	R260 625.87	R286 780.62	R315 858.90	R348 382.27
	Increase in number of animals	1	1	1	1	1
Sheep	Number of animals	249	305	364	425	503
	Old animals sold	25	31	36	42	50
	Auction price	R615.00	R676.50	R744.15	R818.57	R900.42
	Expenses	R411.97	R453.16	R498.48	R498.48	R498.48
	Replacement cost	R1500.00	R1650.00	R1815.00	R1815.00	R1815.00
	Total income	R243 451.52	R328 410.99	R430 592.47	R552 985.71	R720 429.38
	Total expenses	R102 579.98	R138 378.25	R181 433.12	R211 822.15	R250 874.31
	Increase in number of animals	81	89	97	121	152

Figure 3: Different feed yield sections of Bloemhoek Farm – applying Step 1 of the extended Farm Site Development Method.

Step 2: Calculate the arable land

According to Step 2, several inputs are required, which are declared in Table 2.

Method step application: The arable land (a) = (total area (564.34) – existing facilities and road areas (39.05)) = 525.29 ha.

Step 3: Calculate the yield potential to generate a facility layout at saturation state

According to Step 3, several inputs are required, which are declared in Table 2.

Method step application: The yield sections are indicated in Figure 3.

Applying the extension for livestock farms:

Yield potential on a livestock farm is calculated by:

$$y_l = \sum_{j=1}^{k} a_j * u_j, \qquad \text{Equation 4}$$

where y_l is the yield potential (LU per year), a is the section size (ha)……(a_1, a_2,…,a_k) (see Table 3 for values) and u is the

section yield (LU/ha/year) ……(u_1, u_2,…,u_k) (see Table 3 for values).

Thus, using Equation 4, yield potential on Bloemhoek Farm, y_l = 438.01 LU per year.

The yield potential could also be expressed as the number of animals of a particular type (L_i). A linear programming model is used to calculate the values for L_i to maximise profits (Z). Therefore:

$$Max\ Z = \sum_{i=1}^{n}(f_i/(f_i+1))Li(Pi\text{-}Ki)$$

$$i \in I = \begin{matrix} Animal\ type\ 1 \\ Animal\ type\ 2 \\ Animal\ type\ n \end{matrix}$$

Constants:

- y_i = yield potential (438.01 LU per year), calculated in Equation 4
- X_i = ratio animals of type i equivalent to 1 LU, where 1 cow = 1 LU, and $i \in I$

  $X_1 = 1$ (cattle)

  $X_2 = 1/8 = 0.125$ (sheep)
- f_i = number of females associated with 1 male according to the breeding technique for animal type i

  $f_1 = 25$

  $f_2 = 150$
- P_i = the income generated from female animal type $i \in I$ *in one year* i.e. (selling price for animal type i)*(offspring/year o_i for animal type i), in rands (ZAR)

  $P_1 = (R4200)(0.67) = R2814$

  $P_2 = (R615)(1.5) = R922.5$
- K_i = the cost to keep animal type $i \in I$ on the farm *for one year.* The cost calculations are based on feed and veterinary costs associated with the specific farm.

  $K_1 = 2370.74$

  $K_2 = 411.97$

Variables:

- L_i = the number of animals of type i to keep on the farm *annually,* where $i \in I$

Subject to:

- $\sum_{i=1}^{n}\frac{Li}{Xi} \leq y_i$ (saturation state LU may not exceed the LU capacity of the farm)

$\sum_{i=1}^{n}\frac{Li}{Xi} \leq 438.01$ (substituting from Equation 4)

- $X_iL_i = X_i+L_{i+1}\ \forall i = 1,2,3 \ldots . . i\text{-}1$ (relationship of different animal types on the farm remains constant)

Solving the linear programming model renders the following results:

$L_1 = 126$ (cattle)

$L_2 = 2496$ (sheep)

$Z = R215\ 143$ (maximum profit at saturation state)

Because L_i specifies the total number of animals of type i, the breeding technique ratio should be used to calculate $L_{i,f}$ (female animals of type i) and $L_{i,m}$ (male animals of type i):

- $L_{i,f} = (f_i/f_i+1)L_i$ ……from Equation 5

  $L_{1,f} = (25/26)126 = 121$ (cows)

  $L_{2,f} = (150/151)2496 = 2479$ (ewes)
- $L_{i,m} = L_i - L_{i,f}$ ……from Equation 6

  $L_{1,m} = 126 - 121 = 5$ (bulls)

  $L_{2,m} = 2496 - 2479 = 17$ (rams)

Step 4: Determine the production requirements and the saturation date based on constraints

According to Step 4, several inputs are required, which are declared in Table 2.

Method step application: There are two types of animals on the farm; both types will not necessarily reach their saturation dates at the same time. The growth of the number of animals on the farm will be constrained by the cash flow of the business and the percentage of profits that will be used to expand the farm. Furthermore, the annual increase in the number of cattle and sheep needs to be proportional to the shortfall when compared to the saturation-state numbers. The cattle and sheep need to reach their saturation date at the same time to optimise veld utilisation. Table 4 presents a partial cash flow budget, indicating the expenses, income, gross profit and annual increase of animal numbers when 40% of the gross profit is applied to growth of livestock numbers.

The saturation date is the date at which the saturation state is achieved, i.e. when 126 cattle and 2496 sheep are grazing at Bloemhoek Farm. As indicated in Figure 4, the saturation date is at the start of the year 2027. The saturation date is based on certain assumptions, such as normal growth and breeding. Under conditions which impede normal growth and breeding, such as drought, the saturation date will have to be re-calculated using different input values. Section 7 argues in favour of an iterative FSDMe to acknowledge dynamic changes within the enterprise context, such as climate conditions. For the purpose of this demonstration, we acknowledge different input parameters by including sensitivity analyses in Step 9. Because Bloemhoek Farm will be growing in livestock numbers, additional resources, utilities, services and structures will be required, and are calculated in the next step.

Step 5: Identify critical RUSS and design criteria

Additional capacity would be required for sheep sleeping camps, feed-processing equipment and livestock-handling facilities. The main concern or design criterion is the cost of expanding the facilities.

Hanekom[17] indicated that existing sheep sleeping camps had a capacity of 3.2 ha. As calculated in Step 3, the flock of sheep will have increased to 2496 at saturation state. Because a sheep requires 0.004 ha sleeping space, the sheep sleeping camps need to increase to a total capacity of 9.98 ha (i.e. 2496*0.004).

At saturation state, the animals will rely less on natural grazing and more on additional feeding to provide adequate nutrition at different production stages. The different mixtures are also commonly referred to as 'lick', with the key ingredient being lucerne. Hanekom[17] suggested mechanisation as a solution, with additional equipment to produce lucerne at Bloemhoek Farm.

Livestock-handling facilities are used for several activities, such as administering medicine, animal branding and selecting animals for auctions. Existing livestock-handling facilities accommodate 105 cattle and 640 sheep. The livestock-handling facilities require a stepwise increase in capacity.

Step 6: Identify and evaluate alternatives for RUSS replacement or extension

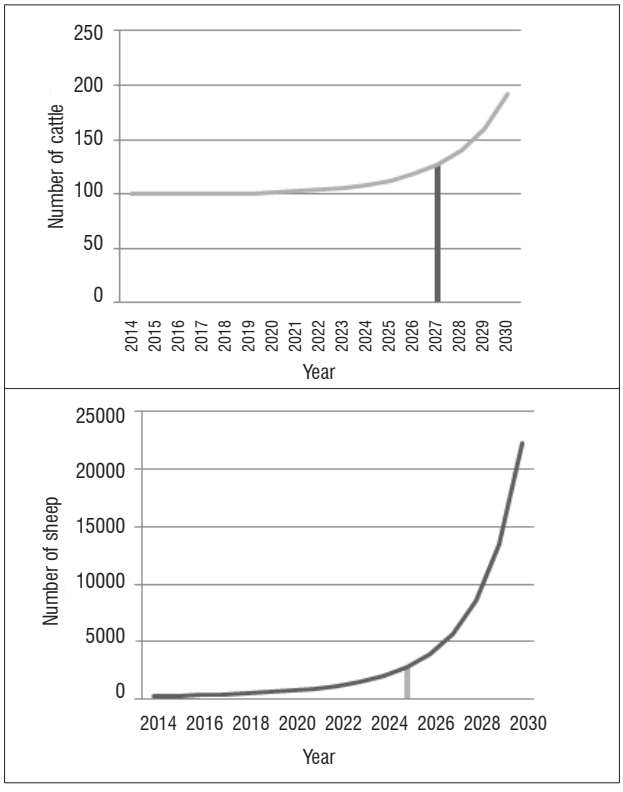
As indicated in the previous step, sheep sleeping camps, feed-processing equipment and livestock-handling facilities have to be extended.

For sheep sleeping camps, the first saturation date will occur on November 2021, as the current sleeping camp has a capacity of 807 sheep (i.e. 3.23 ha/0.004 ha space required per sheep). As indicated in Step 5, a total capacity of 9.98 ha would be required at saturation state. We propose two alternatives: (1) constructing all of the camps before November 2021 and (2) constructing camps as sleeping capacity runs out. The economic analysis indicates preference for the first alternative, as the more economical option.

Regarding feed-processing equipment, Hanekom[17] suggested that the high cost and 100% utilisation of current manual labour warrants partial mechanisation. The suggested lucerne production equipment, detailed in Hanekom[17] (costing R603 042), should be purchased immediately.

Table 4: Effect of animal growth choices on cash flow for Bloemhoek Farm

	Year	2014	2015	2016	2017	2018
Cattle	Number of animals	100	100	100	100	100
	Old animals sold	1	1	1	1	1
	Auction price	R4200.00	R4620.00	R5082.00	R5590.20	R6149.22
	Expenses	R2370.74	R2607.82	R2868.60	R3155.46	R3471.00
	Replacement cost	R10 000.00	R11 000.00	R12 100.00	R13 310.00	R14 641.00
	Total income	R274 433.33	R301 696.31	R331 972.63	R365 633.18	R403 281.71
	Total expenses	R237 074.25	R260 625.87	R286 780.62	R315 858.90	R348 382.27
	Increase in number of animals	1	1	1	1	1
Sheep	Number of animals	249	305	364	425	503
	Old animals sold	25	31	36	42	50
	Auction price	R615.00	R676.50	R744.15	R818.57	R900.42
	Expenses	R411.97	R453.16	R498.48	R498.48	R498.48
	Replacement cost	R1500.00	R1650.00	R1815.00	R1815.00	R1815.00
	Total income	R243 451.52	R328 410.99	R430 592.47	R552 985.71	R720 429.38
	Total expenses	R102 579.98	R138 378.25	R181 433.12	R211 822.15	R250 874.31
	Increase in number of animals	81	89	97	121	152



The vertical line indicates the saturation date.

Figure 4: Growth of livestock constrained by cash flow.

Table 5: Farm site development plan

Year	Increase in cattle	Increase in sheep	Additional facilities	Cost
2014	1	81	Feed-processing equipment (November)	R603 042
2015	1	89		
2016	1	97		
2017	1	121		
2018	1	152		
2019	2	195	Livestock-handling facilities (April)	R129 817
2020	2	255		
2021	2	340	Sleeping camps (November)	R261 360
2022	3	465		
2023	4	652		
2024	5	942		
2025	7	1405		
2026	10	2168		
2027	14	3474		

As calculated in Step 5, the existing livestock-handling facilities accommodate 105 cattle and 640 sheep. Livestock-handling facilities will be depleted in July 2022 for cattle and April 2019 for sheep. Two alternatives are suggested: (1) constructing additional units before April 2019 and (2) constructing the facilities when capacity runs out. The economic analysis indicates preference for the first alternative, based on cost.

Step 7: Compile a series of phase plans: The FDP

The suggested FDP is shown in Table 5.

Step 8: Represent phase plans graphically in support of the FDP

Hanekom[17] provided a graphical saturation-state facility plan for the livestock-handling facilities, as well as the sheep sleeping camps.

Step 9: Validate the FDP

Hanekom[17] automated the linear program and the resulting cash flow using MS Excel. We then used different input parameters to analyse the effect on the FDP. Examples of different scenarios include a decline in the auction price of animals, above average increases in input cost (such as labour cost and feed cost), and a sudden decline in animal numbers as a result of a lethal disease. For demonstration purposes, we demonstrate two scenarios.

Scenario 1: The auction price dramatically increases as a result of El Ninõ, which causes a drought. Typical unit auction prices quoted in February 2016 for cattle were R6000 (instead of R4200) and for sheep R1050 (instead of R615). The result of this scenario is shown in Table 6. The optimal quantities for cattle (405) and sheep (264) at saturation state are very different to the optimal quantities calculated in Step 3. Thus the FSDMe should not be based on volatile prices, i.e. stable price trends should be used.

Table 6: Scenario 1: Auction prices increase dramatically

Higher auction prices as input		Optimal quantities	
Cattle	R6000	L_1 (Cattle at saturation state)	405
Sheep	R1050	L_2 (Sheep at saturation state)	264
		Z (Maximum profit)	R706 021.13

Scenario 2: Above average additional costs are used, e.g. additional costs for cattle are R1000 (instead of R515) and for sheep R300 (instead of R180). The result of this scenario is shown in Table 7. Higher additional costs do not have an impact on the optimal quantities for cattle (126) and sheep (2496), as the same optimal values were obtained in Step 3. Because of higher input costs, the maximum profits are lower than those obtained in Step 3, as expected. In addition, a qualitative validation was performed, using a questionnaire to validate the usefulness of the FSDMe to the farm owner.

Table 7: Scenario 2: Above average additional costs are used

Higher additional costs as input		Optimal quantities	
Cattle	R1000	L_1 (Cattle at saturation state)	126
Sheep	R300	L_2 (Sheep at saturation state)	2496
		Z (Maximum profit)	R116 591.65

Discussion

In this article, we have suggested the extension of an existing method – the FSDM – towards an FSDMe. We have presented the FSDMe, which was adapted and extended to address several utility deficiencies, which are now discussed. The *validity*/*credibility* was improved by extending the FSDMe for livestock farms and incorporating optimisation techniques, demand planning and financial planning. Even though we incorporated demand planning theoretically, our demonstration at Bloemhoek Farm assumed that demand exceeded supply. Additional demonstrations of the FSDMe, in which demand is constrained, would further increase the validity/credibility of the FSDMe. In addition, the purpose/utility statement and scope of the FSDM was clarified by defining the purpose, scope, prerequisites, input requirements and roles. The new FSDMe provides clarity on the structure of the method, separating the initial input variables from the method guidance. The *generalisability* of the method was improved through the distinction between the generic and extension components, as illustrated on crop and livestock farms. The *objectivity*/*confirmability* of the method was improved by involving two objective participants as practitioners, demonstrating the FSDMe for different types of farms.

The two demonstrations partially confirmed the utility of the FSDMe, as a component could not be demonstrated if its method condition did not occur. As an example, Step 4 requires a time series analysis to project future demand *if* market demand is a constraint; because market demand for produce was not a constraint for the two demonstration farms, the time series analysis component could not be validated.

The FSDMe demonstration on the livestock farm acknowledged the heterogeneous nature of the arable land, different feeding habits of animal types and grazing sequence to ensure optimal utilisation of plant material. The FSDMe demonstration on the crop-producing farm acknowledged intervention decisions of the farm owner (e.g. the decision to buy hail shields), which constrains the solution space for developing the FDP.

Conclusions

We have addressed several utility deficiencies of an existing method (the FSDM), extending its *utility* in terms of its validity, its purpose statement and scope, its generalisability and confirmability. In terms of its *purpose*, the extended FSDMe is a strategic planning tool that facilitates production growth on an existing farm of fixed size and incremental extension of its facilities towards the saturation state. The *scope* of the FSDMe includes crop-producing farms as well as livestock farms. For the crop-producing farm, the FSDMe requires choice of crop products as input, but does not aim to optimise the production mix proportions. For the livestock farm, the FSDMe similarly requires a choice of animal type as input, but also optimises the proportion of animal types. Furthermore, it is assumed that the livestock farm produces crops for the sole purpose of feeding livestock. As indicated by the FSDMe, the practitioner or farm owner needs to have some practical knowledge of the farming industry and the farm produce prior to using the FSDMe. Several input values regarding the particular farm, produce and budget are required prior to applying the FSDMe.

The two distinct demonstrations of the FSDMe for a crop-producing farm and a livestock farm validated the utility of the FSDMe. For demonstration purposes, only the livestock farm was included as an example. Feedback from the research participants indicated several extension opportunities for future work. Currently the FSDMe does not question decisions regarding farming produce (i.e. crop varieties, crop production mix and animal types). Because the initial business model may not be optimal, the FSDMe could be extended to re-visit and guide business model decisions regarding markets, produce, partners and resources. One of the assumptions of the FSDMe, when applied to a livestock farm, is that calves and lambs do not graze and are sold once they are weaned. Yet, if the price of crops (e.g. maize) is very low, some farmers rather use a percentage of their crops as rounding feed for calves and lambs. A possible extension of the FSDMe is to incorporate the rounding feed as an input, which will affect the saturation state. Other constraints that need to be considered include crop rotation, land rehabilitation practices and legislative constraints. The FSDMe could also be converted to an iterative method to acknowledge dynamic changes within the enterprise context, such as labour strikes and climate conditions, as well as disasters, such as veld fires and droughts.

Design science research acknowledges that additional research cycles will increase the rigour of the artefact, i.e. the FSDMe, as well as the utility of the accompanying software application that provides a user-friendly interface to the user of the FSDMe.

Acknowledgements

We are grateful that Bloemhoek Farm could be used to demonstrate the FSDMe, and we thank Franko Botes for his assistance.

Authors' contributions

E.H. was a co-designer of the FSDMe and applied the FSDMe to a livestock farm. A.L. was one of the initial designers of the FSDM, critically evaluated the extensions to the original method, and contributed towards the development of the FSDMe. M.d.V. coordinated the project and ensured that the participants adhered to Offerman et al.'s[10] guidelines for good method design.

References

1. Bureau for Food and Agricultural Policy (BFAP). 10th BFAP baseline: Agricultural outlook 2013–2022 [document on the Internet]. c2013 [cited 2014 Apr 10]. Available from: http://www.bfap.co.za/documents/baselines/BFAP_Baseline_2013.pdf
2. Directorate: Statistics and Economic Analysis. Economic review of the South African Agriculture 2013. Pretoria: Department of Agriculture, Forestry and Fisheries; 2014. Available from: http://www.senwes.co.za/Files/main_productsservices/agriservices/2014/Economic_Review_2013.pdf
3. Van Rooyen L. Climate change putting farmers out of business. Farmer's Weekly. 2011 August 23. Available from: http://www.farmersweekly.co.za/article.aspx?id=10388&h=Climate-change-putting-farmers-out-of-business.
4. Seeds of change. The Economist. 2013 June 20. Available from: http://www.economist.com/blogs/baobab/2013/06/land-reform-south-africa
5. Du Plessis D. Reforming land reform. Farmer's Weekly. 2014 February 03. Available from: http://www.farmersweekly.co.za/article.aspx?id=53049&h=Reforming-land-reform.
6. Van der Merwe R, Liebenberg A, De Vries M. A farm site development method: Creating a roadmap towards site saturation. S Afr J of Ind Eng. 2014;25(3):126–139.
7. Harmsen F, Saeki M. Comparison of four method engineering languages. In: Brinkkemper S, Lyytinen K, Welke RJ, editors. Proceedings of the IFIP TC8, WG8.1/8.2 Working Conference on Method Engineering: Principles of Method Construction and Tool Support. London: Chapman & Hall; 1996. p. 209–231.
8. Welke RJ, Kumar K. Method engineering: A proposal for situation-specific methodology construction. In: Cotterman S, editor. Systems analysis and design: A research agenda. Chichester: Wiley; 1992. p. 257–268.
9. Bucher T, Klesse M, Kurpjuweit S, Winter R. Situational method engineering. In: Ralyté J, Brinkkemper S, Henderson-Sellers B, editors. Situational method engineering: Fundamentals and experiences. Boston, MA: Springer; 2007. p. 33–48. http://dx.doi.org/10.1007/978-0-387-73947-2_5
10. Offermann P, Blom S, Bub U. Proposal for components of method design theories: Increasing utility of method design artefacts. Bus Inf Syst Eng. 2010;2(5):295–304. http://dx.doi.org/10.1007/s12599-010-0120-x
11. Gregor S, Jones D. The anatomy of a design theory. J Assoc Inf Syst. 2007;8(5):312–335.
12. Hodson WK. Maynard's industrial engineering handbook. 4th ed. New York: McGraw-Hill; 1992.
13. Jacobs RF, Chase RB. Operations and supply chain management. 13th ed. New York: McGraw-Hill; 2011.
14. Seal W, Garrison RH, Noreen EW. Management accounting. 4th ed. London: McGraw-Hill; 2012.
15. Gregor S, Hevner A. Positioning and presenting design science research for maximum impact. MIS Quarterly. 2013;37(2):337–355.
16. Vaishnavi V, Kuechler W. Design research in information systems [document on the Internet]. c2004 [cited 2009 Dec 11]. Available from: http://desrist.org/design-research-in-information-systems
17. Hanekom E. Farm site development method at a livestock farm [thesis]. Pretoria: University of Pretoria; 2014.