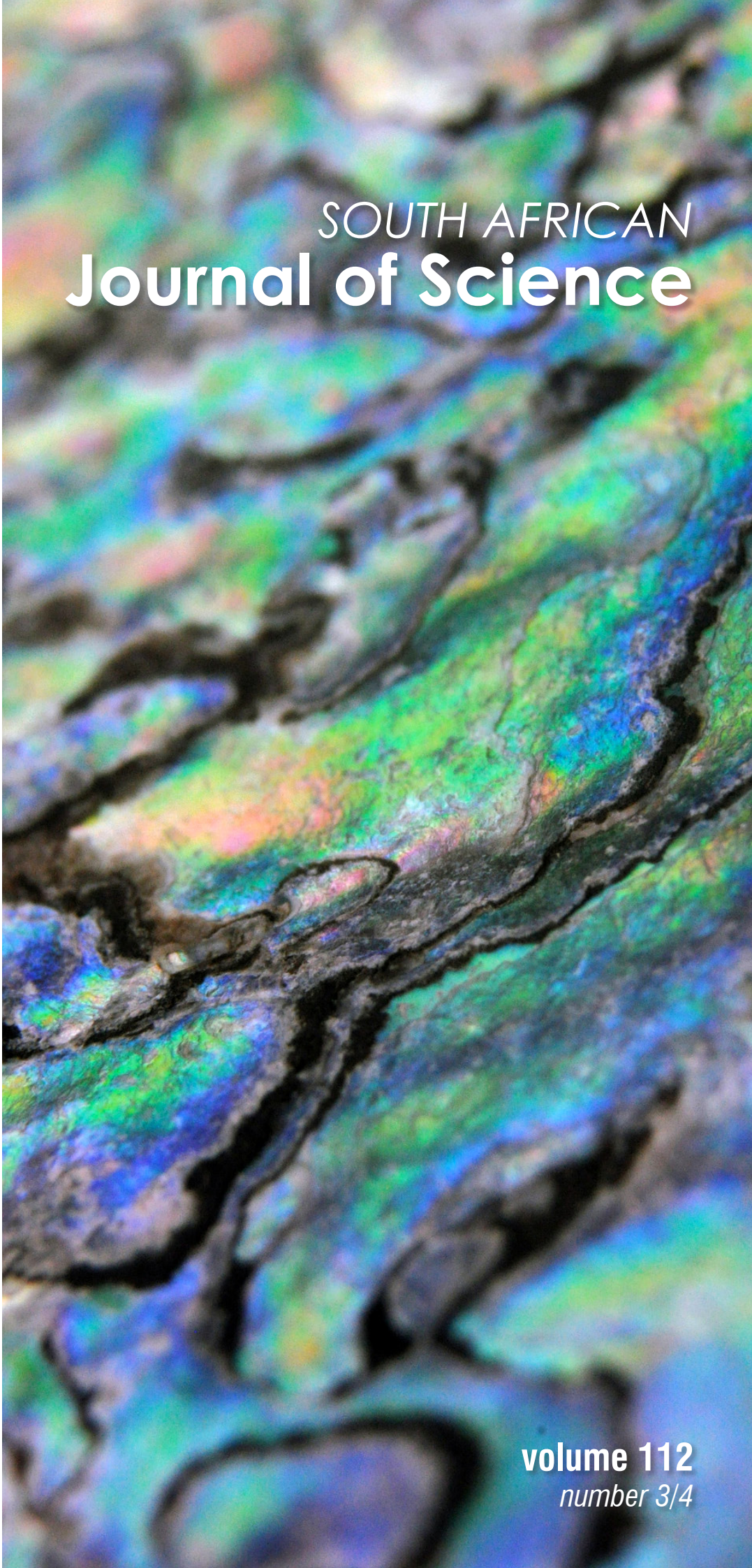
Climate trends in Namaqualand – a biodiversity hotspot

Trading on extinction: an open-access deterrence model for the South African abalone fishery

Antimutagenic and antioxidant effects of a traditional immune booster

Review of treatments for brewery wastewater in a water-scarce country

The undergraduate system in South Africa must be restructured for sustainable funding and fees



SOUTH AFRICAN Journal of Science

MARCH/APRIL 2016

eISSN: 1996-7489

Applying scientific thinking in the service of society

20 YEARS

ASSAf ACADEMY OF SCIENCE OF SOUTH AFRICA

volume 112
number 3/4

MARCH/APRIL

eISSN: 1996-7489

SOUTH AFRICAN

Journal of Science

volume 112

number 3/4

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

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

Correspondence and enquiries
sajs@assaf.org.za

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

Cover caption
An abalone shell (photo: Paxson Woelber / CC BY). In an article on page 105, Crookes shows that a legal trade in abalone in South Africa results in a 'trading on extinction' resource trap, with a focus on profit and a depletion in abalone populations.

Research Article

Research Letter

What really matters for students in South African higher education?

> *He sat at his desk for a few minutes, then drew some papers towards him and looked at them, reading through the words, understanding each one but not certain he understood what they intended to say. He set them aside, then pulled them back to read them again; this time, the sentences made sense to him, though he could see no reason why anyone would find their messages important.*[1(p.49)]

Almost everyone who works in a university or research organisation has had this experience, and not only when reading institutional circulars or exam papers. But the experience has almost certainly increased exponentially in the face of what has been written about '#Rhodesmustfall', '#feesmustfall' and '#outsourcingmustfall' over the past five or six months – along with the volume of analyses of these events, and of what other people have written about these movements. Yet at least three messages that emerge from the endless sentences *are* important, in this instance, as preludes to what really matters for students in the South African higher education system.

Firstly, there is the matter of how to sustain the financing of higher education institutions, without fee increases (and in some cases, without registration fees too) being levied and paid. Assuming that universities should, at the very least, be no worse off than they were in 2015 in view of the rising consumer price index – which applies to operating costs and to acquiring intellectual materials with a weak Rand – the value of the unpaid increases needs to be sourced. There is a scramble underway to attempt to ensure that this happens. At the same time, students (both *revenant* and new) will need support for the fees that do still have to be paid; for their learning materials; and, critically, for their living costs.

For a very few, this will come from their families, but for students from 'squeezed' middle class families, and from families who are unable to make any contribution to these expenses (taking account of the unemployment figure of 8.3 million South Africans – affecting close to half the population), the National Student Financial Aid Scheme (NSFAS) will have to do better. It will need more funding, and it will need to be better administrated, both within the Fund and within universities. The appointment of Sizwe Nxasana as the new Chair of NSFAS is a remarkably positive step – but turning the Fund around may prove to be a task at least as challenging as managing First Rand.

The second issue spirals around the debates regarding free education for all. This shift in the '#feesmustfall' campaign was sparked by President Zuma's announcement last October that fees for 2016 would be capped at 2015 levels, stating that the government was understanding of the difficulties faced by students from poor households – and urged all students to allow the process to unfold to find long-term solutions. In response to this, he set up a Commission of Enquiry, none of whose members are educationists or economists, to investigate whether or how free higher education might be implemented for higher education students. Notably however, despite the costs of the student protests – damage to university buildings, deferred examinations, and problematic registration processes this year – the President devoted no more than two or three sentences in his State of the Nation Address to higher education. It is perhaps fitting then, that one of the banners carried by students protesting outside the Union Buildings last year declared 'Time For Educated Leaders!'

Articles in the popular press by those who have addressed this issue range from those of serious economists (including Thomas Picketty) who indicate that free education for all is primarily a benefit to the wealthy; through to the damage that a state of financial scarcity does to sound educational decision-making; to the position taken by some who ignore economic reality and government expenditure patterns in order to claim that free higher education for all is undoubtedly possible; on to those who point to the failure of free higher education in other African countries. But in the face of South Africa's budget deficit and the fear of a junk bond status (and so higher borrowing costs), this seems to be a poorly considered position – the more so if the greatest benefits accrue not to the students who most need free education but to the economically better off.

The '#Rhodesmustfall' movement, of course, was the public face of the third issue (and another, addressed below): the decolonisation and (or) the inclusion of indigenous knowledge in the curricula of all universities. Clearly, both the 'decolonisation' of knowledge and respect for, and the inclusion of, relevant indigenous knowledge in the curriculum, are issues of the greatest importance. They are not, however, unique to South Africa, nor – in many instances – are they new issues. Many departments of literature in South African universities have, for instance, been teaching African literature and not only on the basis of 'European' theories of analysis. This might not yet be true in all departments of philosophy, sociology, or history, for instance (although there are some examples), but the idea is not entirely new. What's more, there are some limits to the process: how might genomics, the general theory of relativity, gravitational waves or the 10-min saliva test for cancer be decolonised – or enriched by indigenous knowledge? This is not to suggest that there are no real issues at stake for curriculum revision, but the demands should not overlook what has been done, what can be learned from those practices, as well as the limits that exist if South African universities are to teach disciplines that are, in fact, respected elements of international research. Relating those respected research findings to local contexts will, of course, require a knowledge and understanding of local circumstances – but the core knowledge will remain as it is. Until South African researchers change it.

The three issues are highly significant. Recognising them, taking them seriously, and dealing with them in ways that are intellectually rigorous and honest, are all essential to the future of higher education – not in South Africa alone, but in many other parts of the global south and north.

Access to higher education – to any level of education, of course – is critical and affordability is too, because access without affordability has no meaning; and do not forget sound and relevant curricula and teaching skills. But what matters after those issues are addressed? What counts next? Probably the most important matters of all are those to which the earlier issues are the preliminaries. Why take the trouble to access higher education, at no or low cost, with changed curricula and teaching (issues that might well be matters for contention), if what is learned is of low quality? If the worst of the hurdles are removed, but the race is not worth the running, it is all to no purpose. So the most critical matter that counts next is the high quality of the content, of the science, and of the research offered by institutions and recognised as such, not just in Europe or the USA, but also, of course, in India, China or Brazil. Adam Habib, Vice Chancellor of the University of the Witwatersrand, and Chair of Universities South Africa, says:

> *There is a danger in this moment that if we allow the current populism to be unconstrained, it could result in a higher education system that enables access, but destroys quality. This is the history of the continent and it would be a tragedy if it were to be repeated. From 2016 onwards it is going to be a political and intellectual struggle between these two outcomes.*[2]

HOW TO CITE: Butler-Adam J. What really matters for students in South African higher education? S Afr J Sci. 2016;112(3/4), Art. #a0151, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0151

Sadly, as this Leader is being written, the populism, partly fuelled by the '#Rhodesmustfall' movement, continues to be widespread, and completely misaligned with the fundamental needs of South African society. There are campus protests at four universities, some involving student arrests by police. The protests at one university included the burning of works of art of historical value and the fire-bombing of the Vice Chancellor's office and, at another, the destruction of a science laboratory for schoolchildren. The wealthy parent of one of the '#Rhodesmustfall' 'activists' involved in the protest, who is the CEO of a state owned enterprise, claimed that this 'builds character' and that it would 'spark intellectual debate'. Really? These are serious, criminal, destructive activities. The hapless indulgence and dismissal of law and order on the part of a parent, however well-placed she or he is, cannot be ignored, because this tolerance contributes, collectively, to the problem of a social movement that serves its own self-interests – and nothing more than that. Yet another parent, who fought consistently for liberation and democracy, is silent regarding his daughter's destruction of public property.

The second condition that matters (if the first were not disheartening enough) is learning that ultimately leads to the opposite, the creation of social value. If earning a worthless degree is a useless exercise, then earning a degree that is of no personal or social use is worthless. This is most definitely not intended to suggest that higher education must be nothing other than an instrumentalist process: being employable means, more than anything, having high level analytical and communication skills, the ability to solve problems, to be innovative, and adaptable. It is not just about learning content, but also about developing flexible talent, and about how to make the most of those skills in new and challenging circumstances. In some instances, problem-solving and adaptability skills are of greater value than content that might have a brief half-life.

The third 'what matters' is learning, and then graduating with, knowledge and values that prepare students to be successful, confident world citizens. This does not imply that graduates should leave South Africa (although they might well choose to do so). More significantly, it points to the fact that South Africa's future depends on being part of a wider world in which countries, including South Africa, are generating new ideas, applications, and economies that are the foundation of our own national survival – and of the world.

There's a great deal at stake, more than can immediately be imagined, for higher education in South Africa, and so also for young South Africans (poor or rich) who enter universities; and even more at stake when it comes to their contributions to their own and the country's success and prosperity. We cannot afford to ignore the immensity of the consequences if we do not do everything possible to get higher education right.

Quidnunc?

References

1. Leon D. Uniform justice. London: Heinemann; 2003.
2. Habib A. Goals and means – Reimagining the South African university. University World News. 2016 Feb, Issue 400. Available from: http://www.universityworldnews.com/article.php?story=20160211070052665

ASSAf 1996–2016
Commemorative article

The essence of scholarship: Charting a path through the thickets of scholarly publishing

AUTHOR:
Robin M. Crewe

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

CORRESPONDING AUTHOR:
Robin Crewe

EMAIL:
Robin.Crewe@up.ac.za

POSTAL ADDRESS:
Centre for the Advancement of Scholarship, University of Pretoria, Private Bag X 20, Hatfield 0028, South Africa

KEYWORDS:
Academy of Science of South Africa; SciELO SA; peer review; open access

HOW TO CITE:
Crewe RM. The essence of scholarship: Charting a path through the thickets of scholarly publishing. S Afr J Sci. 2016;112(3/4), Art. #a0148, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0148

The history of scholarly journal publishing is generally dated from the appearance of the *Philosophical Transactions of the Royal Society of London* in 1665.[1] The notable features of this publication are that it is the property of a learned society, and at the time of its inception reflected the deliberations of the scholars of the day from across Europe. It gained its stature from contributions by members of the learned society and the esteem of a broader scholarly community that used it as a vehicle to exchange ideas. Since the inception of the notion of the 'scholarly journal', the number of journals, and the range of subjects covered, has proliferated dramatically. In addition, the model for journal publishing has moved from the historical learned society publisher through to the emergence of large commercial publishers who dominate the market. The rise of electronic publishing has made it possible for predatory publishers with no pretensions of quality to join the fray. The various journal databases currently list in excess of 30 000 reputable titles and with the emergence of open-access online journals, the number is being proliferated extravagantly and with little regard for the quality of what is published. For an aspiring scholar who is looking for a credible vehicle in which to publish his or her work, the choices are bewildering. But informed choices are crucial for establishing a scholarly reputation. Thankfully, help is at hand in the form of the work that has been done in South Africa to enhance the reputation of local scholarly journals.

In South Africa, a variety of initiatives has been launched in the past by groups of academics and learned societies to establish journals as vehicles for scholarly communication. Indeed, there was even a government initiative that started in the 1970s to provide an infrastructure to support selected South African journals and enhance their impact in the global scholarly community.[2] With the demise of these initiatives in the 1990s, the local scholarly publishing landscape appeared drought-stricken until the turn of the 21st century.

However, two local interventions have played a crucial role in the lives of scholars who are intent on having their work published and establishing their reputations. The first of these was the introduction by the Foundation for Research Development (ancestor of the current National Research Foundation (NRF)) of a rating system for individual scholars in 1985. The rating of individuals was based on peer review of their scholarly contributions to their disciplines. The ratings essentially assessed whether they were recognised by their peers as falling in the broad categories of being international leaders in their fields, being recognised internationally for their contributions, or being recognised nationally. This assessment was based on the evaluation of the significance of a person's particular contributions to the scholarly literature and was influenced by the quality of journals in which the work appeared. The ratings given to scholars became an important factor in the development of academic careers, particularly when scholars in the humanities and the social sciences were included in these ratings.

The second intervention was the change to a funding framework for universities in 2003[3] that provided an output subsidy for scholarly publications in journals, conference proceedings and books. For universities, maximising the number of these outputs was an important source of income, while for the government department providing the funding, this was meant as an incentive to enhance research performance, but with quality criteria built into the recognition of these outputs.

An aspiring scholar in 2004 was confronted with competing demands of the NRF for quality of scholarly work related to high ratings, and their institution's demands, both for an NRF rating and a greater number of outputs to enhance income through the research output subsidy. This latter problem of numbers could be partially finessed using South African journals that had a special position as recognised journals for subsidy purposes. The use of 'in-house' journals for this purpose was clearly an attractive option to pursue if the level of NRF rating could be traded off against income generation.

First report on scholarly publishing in South Africa

Into this conflicted terrain of scholarly publishing politics stepped the then newly established Academy of Science of South Africa that was asked by the then Department of Arts, Culture, Science and Technology (DACST) to initiate a study of research publishing in South Africa. The impetus for this request came from the understanding at the time that roughly half of the research outputs from South Africa came from publications that were in Web of Science listed journals and the other half were in journals that were not listed by Web of Science but recognised by the Department of Education. Indeed, 219 South African journals were recognised by the Department of Education in 2004. The journals were diverse and it appeared that 'their primary purpose may not be communication and documentation of original research in a global knowledge system.[4] In view of these reservations about the quality of local journals, the DACST requested that the Academy carefully examine the evidence available regarding South African research journals and develop a new strategic framework that would be comparable with the situation prevailing elsewhere in the global academic environment. This DACST contract was to be a profound test of the young Academy's ability to undertake a thorough investigation that would lead to implementable policy recommendations.

The report took as its starting point the identification of key properties of a research journal that would provide a reliable record of new knowledge being added to the global corpus of scholarly knowledge. The authors of the report identified three essential characteristics that all credible research journals needed to exhibit. The first was that readers should be able to place an absolute reliance on the integrity of the research results being presented, in terms of both methodology and interpretation. The second was the core role of the editor in managing the evaluation of submitted manuscripts and the peer-review system associated with their evaluation. Finally, the authors of the report recognised that the nature of scholarly publishing was changing radically and that the

electronic dissemination of research information was changing the nature of the scholarly enterprise in ways that were evolving rapidly and needed to be assessed. In this latter respect, the report was particularly prescient when it was initiated in 2001.

The work required for careful examination of the evidence was concluded in 2005 and the report was published in 2006. This report provided an incisive analysis of scholarly publishing and remains an extremely useful source of information to guide both authors and journal publishers in carrying out their respective roles responsibly. For example, there is a very useful definition of a South African journal[1(p.2)] in order to avoid arguments about what 'South African' journals are. In addition, the analysis of South African journals at that time provided a unique insight into the state of local scholarly publishing, not all of it very flattering. As with all reports of this kind, it provided a range of recommendations for the appropriate government departments to consider in the development of their policies and for the sector to consider as journals and their editors grappled with the findings.

Impact of the report

In the decade since the report was published, what impacts have the recommendations had?

All ten of the recommendations have been largely implemented with the Department of Science and Technology and the Department of Higher Education and Training (DHET) partnering with the Academy to achieve the vision set out in the original report, of establishing a vibrant local scholarly publishing environment that engages globally in making South African knowledge generation visible. In addition, there has been an attempt to ensure that knowledge generated locally is made accessible to learners in schools so that they appreciate that knowledge generation is an indigenous activity in which they can become active participants.

In taking on the task of implementing the recommendations of the report[1], the Academy:

- Undertook a study of scholarly book publishing[5] to complement the work on journal publishing. This report made a number of recommendations, some of which have recently been incorporated into the new guidelines[6] of the DHET for research output recognition that apply from 2016. This study was also definitive in dealing with the elements of what constitutes a scholarly book and how these should be evaluated for the purposes of the output subsidy.
- Established the Committee on Scholarly Publishing in South Africa that advises the Council of the Academy on matters related to scholarly publishing and oversees the activities of the administrative unit within the secretariat of the Academy that is known as the Scholarly Publishing Programme.
- Established the open-access journal platform SciELO SA that is designed to be the premier collection of research journals from South Africa. From its establishment with the *South Africa Journal of Science* as its first journal to the current time with a collection of 60 titles, it has proved to be an excellent platform for enhancing the global visibility of research reported in these titles. Site visits have gone from 5000 in 2009 to 1.3 million in 2015. In addition, the SciELO SA platform has been included in the Web of Science Portal to allow for enhanced searching of the material in the collection.
- Transformed the *South African Journal of Science* into a fully open-access journal and the first to be available on the SciELO SA platform. The Academy took the bold step of pioneering the publication of this fully open-access journal to serve as a model to be emulated by other South African journals in the future.
- Publishes the magazine *Quest* as a means of making South African research activities accessible to a broader audience with the intention of luring young learners into research careers.
- Established and maintains the National Scholarly Editors' Forum that provides a platform for the editors of scholarly journals to get together and consider matters that need to be addressed in relation to the publishing environment in South Africa.
- Established and maintains the National Scholarly Book Publishers' Forum that provides a site for local book publishers to deal with matters of common interest.
- Established and continues to undertake a systematic discipline-based peer review of journals that are published in South Africa. Journals that are approved by this peer-review process may be asked to join the SciELO SA platform and the articles published in them are eligible for the output subsidy.
- Established and undertook a review of submissions from higher education institutions of books, chapters in books and conference proceedings in order to make recommendations to the DHET regarding awarding of subsidies. In this respect the Academy ensured that the assessment of submissions was undertaken by specialists within the disciplines of the authors and established what is considered to be a credible process of assessment.

This constellation of initiatives by the Academy provides for a rich environment in which scholarly publishing in a variety of modes can be pursued. The key elements of these initiatives are to ensure that scholars locally have a variety of vehicles through which to make their work known, to ensure that the quality of the work that is published is maintained at a high standard, both through the peer-review process for individual submissions as well as through the discipline-based peer review of groups of journals themselves, and to provide a platform for global visibility.

For the aspiring scholar of 2016, the demands that they face to obtain ratings from the NRF and to publish regularly as required by their institutions remain the same as for their predecessors at the turn of the century. However, the milieu in which they undertake their work has changed almost beyond recognition through the pervasive use of electronic means to communicate information and ideas, and the changing nature of publishing. They are beneficiaries of comprehensive interventions by the Academy to try to ensure that the quality of scholarly publication is maintained, but they have also been provided with an internationally recognised platform for the dissemination of work published in local journals. Apart from Brazil, South Africa is probably one of the few countries in which such a comprehensive set of interventions has been attempted in support of its scholarly community.

The other lesson of particular significance for the Academy is that the methodology employed to produce the two reports on scholarly publishing[1,6] has been singularly successful in showing what can be produced by studies that have insightful analyses of the evidence, coupled with practical guides to policy development and implementation. The Academy has shown itself to be particularly adept at the implementation of the recommendations of the reports with the provision of a set of interventions that support scholarly activities.

References

1. Academy of Science of South Africa (ASSAf). Report on a strategic approach to research publishing in South Africa. Pretoria: ASSAf; 2006. Available from: http://www.assaf.org.za/files/2011/02/2466-ASSAF-Strategic-approach-to-research-publishing-2.pdf
2. Baker G. Scholarly publishing in South Africa: Facing reality. S Afr J Sci. 2008;104:411–412.
3. Department of Education. Policy and procedures for the measurement of research output of public higher education institutions. Pretoria: Department of Education; 2003. Available from: http://www.dhet.gov.za/HED%20Policies/Policy%20for%20Measurement%20of%20Research%20Output%20of%20Public%20Higher%20Education%20Institutions.pdf.
4. Gevers W. Introduction and background. In: Report on a strategic approach to research publishing in South Africa. Pretoria: ASSAf; 2006. p. 9–28. Available from: http://www.assaf.org.za/wp-content/uploads/2011/02/2466-ASSAF-Strategic-approach-to-research-publishing-2.pdf.
5. Academy of Science of South Africa (ASSAf). Report on scholarly books: Their production, use and evaluation in South Africa today. Pretoria: ASSAf; 2009. Available from: http://www.assaf.org.za/ASSAF%20Scholarly%20Report%20FINAL%20Proof.pdf
6. Department of Higher Education and Training (DHET). Research outputs policy, 2015. Government Gazette. 2015 March 11. Vol. 597: No. 3855215.

For sustainable funding and fees, the undergraduate system in South Africa must be restructured

AUTHOR:
Nico Cloete[1,2,3,4,5]

EMAIL:
ncloete@chet.org.za

AFFILIATIONS:
[1]DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, Stellenbosch, South Africa
[2]Higher Education Research and Advocacy Network in Africa, Stellenbosch, South Africa
[3]Department of Education, University of Oslo, Oslo, Norway
[4]Centre for Higher Education Trust, Cape Town, South Africa
[5]Institute for Post-School Studies, University of the Western Cape, Cape Town, South Africa

POSTAL ADDRESS:
CHET, PO Box 18094, Wynberg 7824, South Africa

KEYWORDS:
Academic ranking; performance; private returns; completion rates; access

HOW TO CITE:
Cloete N. For sustainable funding and fees, the undergraduate system in South Africa must be restructured. S Afr J Sci. 2016;112(3/4), Art. #a0146, 5 pages. http://dx.doi.org/10.17159/sajs.2016/a0146

South Africa has the most diverse and differentiated higher education system in Africa – despite some persistent attempts at academic drift and mimetic normative isomorphism. Globally, in the 2008 country system ranking by the Shanghai JiaoTong Academic Ranking of World Universities, the South African higher education system was placed in the range between 27 and 33 along with the Czech Republic, Hong Kong, New Zealand and Ireland. It is well known that South Africa consistently has four of the five African universities that appear in the Shanghai top 500.

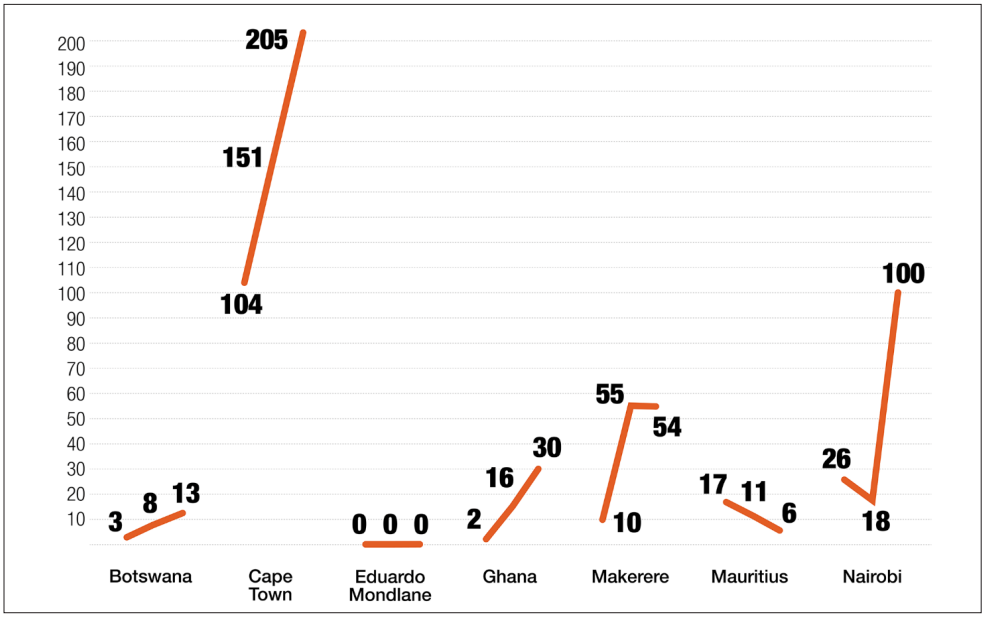
Even more impressive is that The Times Higher Education 2016 ranking of BRICS and emerging economies[1] places three South African universities in the top 12: the University of Cape Town (UCT) 4th, the University of the Witwatersrand 6th and Stellenbosch University 11th. Brazil and Russia each have only *one* university in the top 12, and India, with a billion people, has none. China, with their differentiation policy aimed at producing 30 world-class universities, has six in the top 12.

In the Higher Education Research and Advocacy Network in Africa research programme, which consists of seven African flagship universities, UCT, the only South African university, published 2390 articles in 2014 journals that are listed in the Web of Science with the other six universities combined publishing only 1476. Similarly, in terms of doctoral production, UCT produced 205 graduates in 2013/2014, while the other six universities combined produced only 207 (Figure 1).

However, it is not only in terms of growth that South African universities have excelled in relation to those in the rest of Africa. There have also been considerable efficiency increases. Figure 2 shows that while the number of academic staff increased by 26%, publication output increased by 150%. The doctoral supervision load increased from 4600 academics supervising 5100 students in 1996 to 6700 academics supervising 13 900 students in 2012. In addition, in terms of years to graduate, South Africa did not perform as well as countries such as Norway, the USA and the United Kingdom, which have large proportions of full-time doctoral students. However, in terms of part-time students, South Africa was comparable to the United Kingdom.[2]

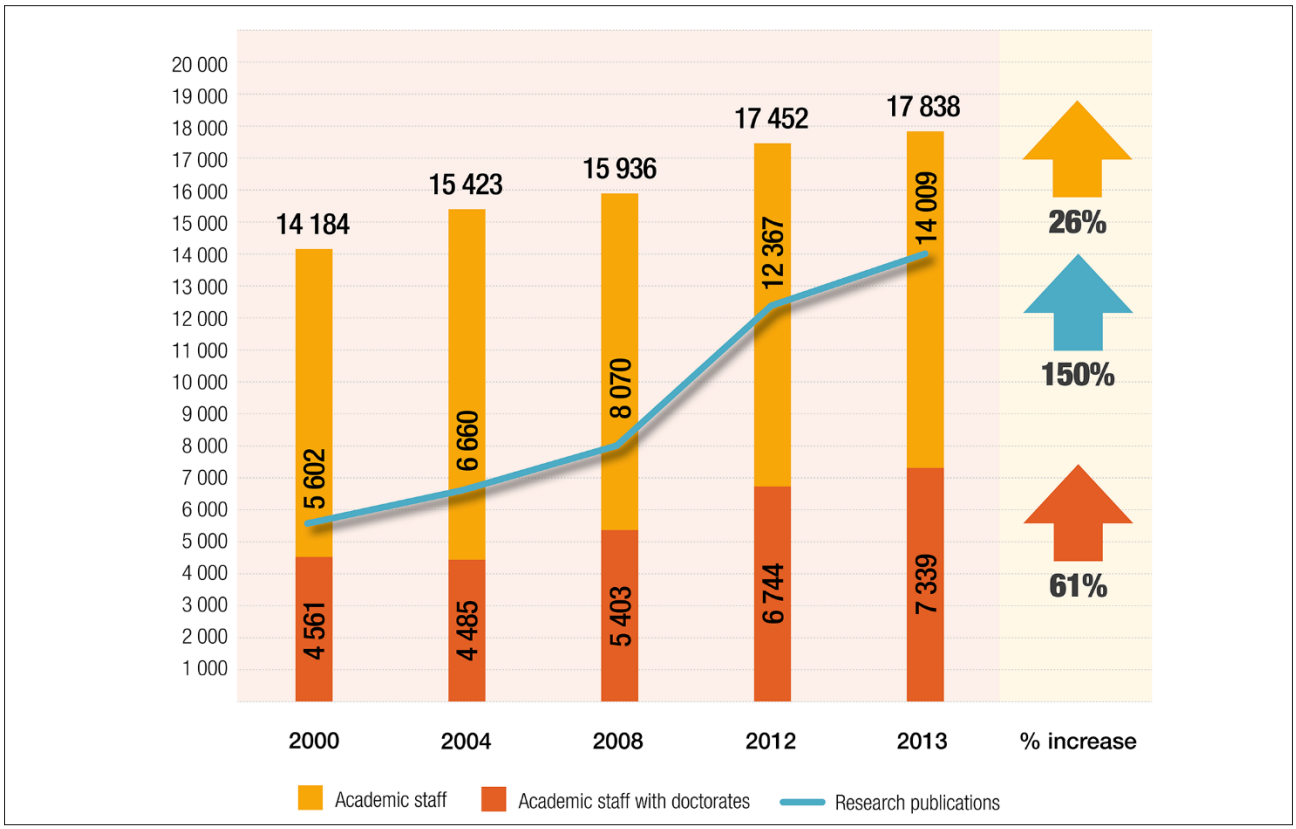
In a book on the doctorate in South Africa, Cloete et al argue that the model of doctoral education requires a radical change that would include moving from 40% full-time students to over 60%, different types of doctoral programmes and full-time students being employed as 'pre-docs', similar to post-doctoral students.[2]

One of the most vocal criticisms against the South African higher education system at the postgraduate (doctoral) level has been the charge of a lack of transformation.[3] The term 'transformation' has become so ideologised that it has little research or policy value. Perhaps one of the most inappropriate ways to use transformation is as a static concept; for example, to demand that universities must reflect, 20 years after apartheid, the demographics of the current population.[4] What we should learn from this charge of a lack of transformation at postgraduate level is that bad policies have long-lasting consequences and cannot be redressed or wished away in a decade or two.



Source: Centre for Higher Education Trust/Higher Education Research and Advocacy Network in Africa data (2015)

Figure 1: Doctoral graduates at seven African universities (2001, 2009, 2014).



Source: Department of Higher Education and Training's Higher Education Management Information System data for 2013. Compiled by Charles Sheppard.

Figure 2: Academic staff and research output at South Africa universities (2000–2013).

However, if transformation is understood as an indicator of change, then South African higher education has undergone seismic changes. Regarding the composition of the entire student body – a largely undergraduate population – Cooper and Subotzky[5] declared that South Africa had experienced a 'revolution' and, by 2013, 74% of all higher education students were black.[6]

Some of the most substantial changes happened at the doctoral level. African doctoral graduates increased from 58 in 1996 to 821 in 2012, an increase of 706% in the post-1996 period. By contrast, white graduate numbers only grew by 71% (from 587 to 816). Over the same period, the proportion of African doctoral graduates increased from 8% to 44%, and in 2012, the number of African graduates exceeded those of whites. African female graduates, starting from a very low base of 10 in 1996, increased by 960% graduates to 106 in 2012, while African male graduates increased by 356%. By contrast, the number of white male graduates remained more or less constant – around 367 between 1996 and 2012. White female graduates increased from 219 in 1996 to 449 in 2012 (105%). If transformation is counted as improvement in percentage change, then Africans (and especially female Africans) have attained spectacular gains, particularly if contrasted to white males. We have not found another international example with such demographic changes in a national higher education system over such a short period (16 years).[2]

And, it should not be forgotten that from 2016, one of the world's largest science projects, the Square Kilometre Array – an international effort to build the world's largest radio telescope, with a square kilometre (one million square metres) of collecting area – will be led by scientists affiliated to South African universities.

One of the factors that sets UCT apart from the other African flagship universities mentioned above is that these flagship universities do not charge fees – they are all part of country systems in which public universities are free and those in private (no research undertaken) are not.

A very inefficient undergraduate system

The 'best' system described above is based mainly on the postgraduate system, which in South Africa is about 16% of the total higher education system,[7] while at certain universities, such as UCT, it is over 30%. In the rest of Africa, the postgraduate systems comprise less than 5% of the total higher education system.[2]

A detailed analysis of the 2000 and 2006 cohorts shows that the proportion of intake into contact institutions of students who are sufficiently prepared to complete undergraduate curricula within the intended time, is small: only 27%, or roughly only one in every four. Performance is very poor for all groups across the three qualification types (diplomas, 3-year and 4-year degrees) with only 48% in contact universities graduating within 5 years. It is estimated that 45% will never graduate. For distance education, the figures for the University of South Africa are simply horrendous. Only 6% of students graduate within 5 years and it is estimated that 78% will never graduate. By the end of the regulation time for all three qualification types, more students have been lost to failure and dropout than have graduated – more than twice as many in the case of African students and those in diploma courses.[8]

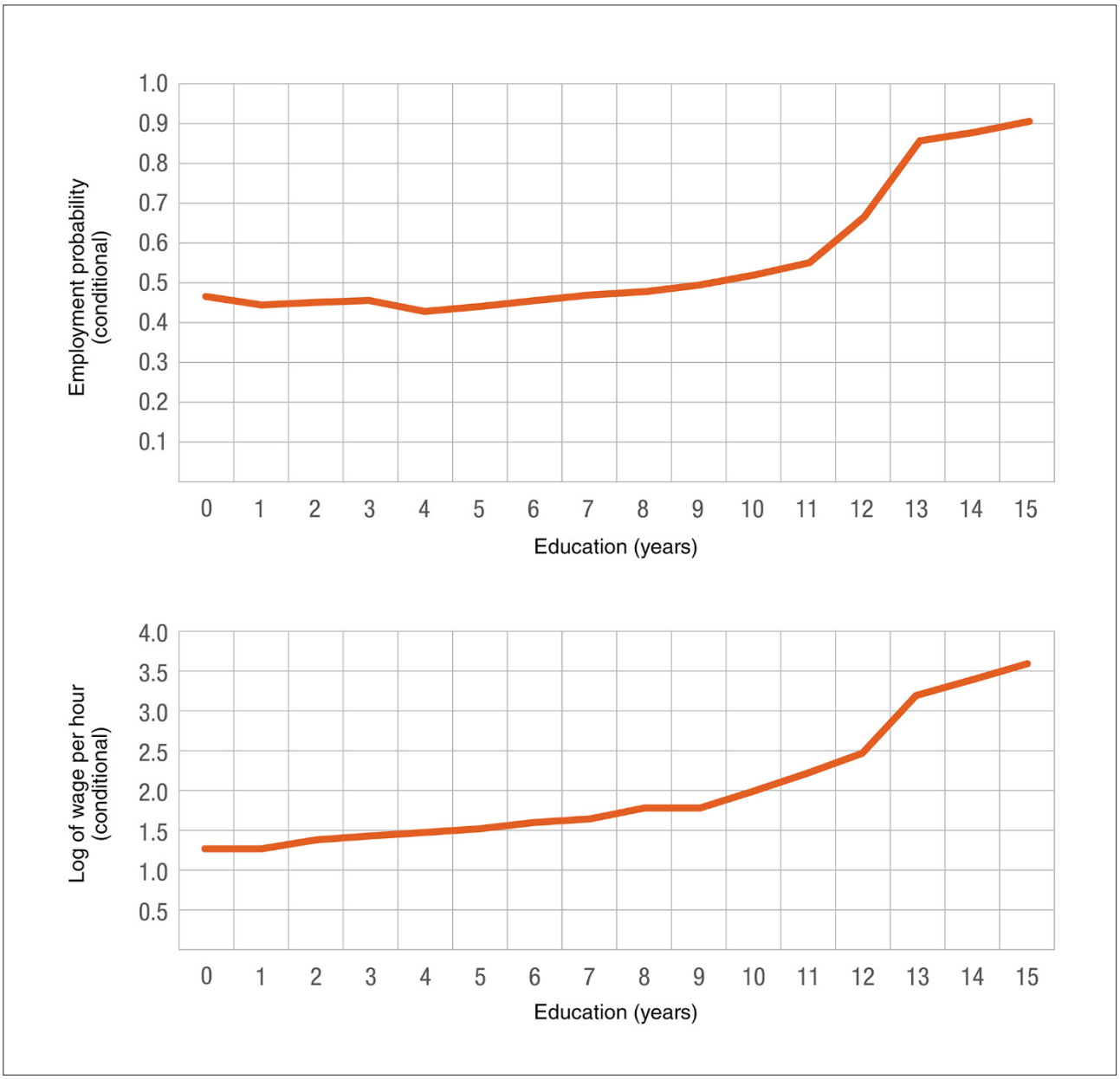
Another method of assessing inefficiency is analysing the total number of undergraduate students entering and exiting the public university system on an annual basis. The Higher Education Management Information System (HEMIS) data show that total undergraduate enrolments in South Africa's public universities grew by 194 000 in 2013 compared to 2006, with less than 10% of the growth among first-time-entering undergraduates. The average annual growth rate for first-time-entering undergraduates between 2006 and 2013 was only 1.7%, compared to an average annual growth rate of 4.7% for the category of undergraduate students who had previously been in the university system. Undergraduate students in South Africa have high dropout and low graduation rates, which result in them remaining registered for long periods, well beyond the normal times required for the completion of their qualifications.

Higher education and inequality

According to both Van den Berg[9] and Piketty[10], technological innovation and globalisation have pushed up demand for highly skilled knowledge workers, even in service industries. If the supply of skills does not increase at the same pace as the growth in technology, then groups whose training is not sufficiently advanced will earn less. Furthermore, with more competition they are relegated to devalued lines of work, which increases inequality. In such economies, the 'haves' are the educated – and the more education, the better – while the 'have-nots' are those who did not finish school or did not graduate from tertiary education. The 'college premium' is known to all, and for many families justifies going deeply into debt to get that degree.[11]

The effect of the kind of university system described above is reflected in a severely distorted labour market and skewed private returns (by students) to tertiary education. Van den Berg[9] found that after controlling for a range of variables such as gender, experience, location, education does bring some rewards. However, the returns below matric are very low. It is only after matric, and particularly at the level of degrees, that returns are extremely high, both in wages per hour but especially in employment probability (Figure 3). Van den Berg's interpretation is that it is only certificates such as a matric (validated by a national exam) and tertiary certificates that signal to employers reliable cognitive gains. Statistically, there are still racial differences 20 years after apartheid, but for Van den Berg these differences are mainly a result of differences in quality of education. He concludes that:

> *The large differentials in earnings and access to jobs between the highly educated and the less educated lies at the heart of income inequality. The high wage premium to educated workers derives from a combination of a skills shortage at the top end of the educational spectrum, driving up wages of the educated, and a surfeit of poorly-educated workers competing for scarce unskilled jobs, thus dampening unskilled wages.*[9(p.214)]



Source: Van den Berg[9]

Figure 3: Conditional probability of employment and conditional log of wages by years of education.

Montenegro and Patrinos[12], in a background paper for the 2013 World Development Report, calculated private rates of returns from 800 household/labour force surveys. This study produced two surprises: firstly, that higher education has higher returns than primary education and, secondly, that the country in the world with the highest private returns to tertiary education is South Africa. The rate of return increased from 28.7 in 2000 to 39.5 in 2011, which is the same period that the Gini coefficient deteriorated from around 0.60 to 0.70.[13] Responding to an email (14 December 2015) that sought to check whether the World Bank finding was correct, Patrinos confirmed the result and stated: 'I believe that high returns to tertiary and high levels of inequality are consistent.'[12]

The returns in South Africa are not just the highest by a small margin; the only other country with a figure over 30 was Rwanda in 2005, but they subsequently improved to 28 in 2010. Ghana and Côte d'Ivoire at 28 have the next highest returns in the world. To illustrate how big the disparity is, these are the figures for a selection of other countries: Mauritius 21, Mexico 20, Brazil 17, Portugal 14, Turkey 14, the USA 14, Argentina 12, Spain 11 and Norway 10.

Access to higher education is regarded by the haves as a means to maintaining privilege and by the have-nots as a means of getting out of poverty. But Piketty points out that in the US, the level of wage inequality results directly from a failure to invest sufficiently in higher education.[11] High tuition at both public and private universities keeps many individuals from receiving the training needed to shrink wage inequality and to make the country more equal and competitive globally. Given such trends, Piketty anticipates that social mobility will decline even further in the future as income increasingly determines access to US higher education. This problem is both amplified and racialised in South Africa: returns to higher education in South Africa are triple that of the USA and as in the USA, are also racially biased.

However, unlike the USA, the South African problem is exacerbated by a low participation rate, low undergraduate completion rates, and the absence of a college sector that can serve as an absorber for poor students, who are also academically and socially underprepared for graduate study. South Africa attempts to maintain a high level of quality, with very high rates of return for a completed undergraduate degree, but then also expects higher education to be a mechanism for reducing inequality. As far as I am aware, there is no system in the world that can achieve such an outcome.

The South African undergraduate system is too expensive, mainly as a result of government underfunding and inefficiencies at the undergraduate level. Thus, it cannot produce large numbers of highly skilled graduates (to drive down the exorbitant rates of return); neither can it absorb large numbers of successful (academically and materially) poor students. As the statistics from the Council on Higher Education[8] show, what the South African undergraduate system is actually doing is taking in large numbers of students who they know have about a 30% chance of completing in 5 years. The universities have been able to maintain this unsustainable system through fee increases and a perverse incentive subsidy system.

Over the last decade, the government subsidy has decreased as a component of total university income from 49% to 40%, while the contribution from student fees has risen from 24% to 31%. It is difficult to gather information on university fees given the variation in costs across degree programmes; however, Statistics South Africa does collect information on higher education course costs from across the country and publishes this information in a 'tertiary education inflation index' annually.[14] This index shows that between 2010 and 2011, the consumer price index was around 5% while the tertiary inflation index was close to 10%. From 2012 to 2014, the consumer price index hovered around 6% while tertiary inflation was between 9% and 10%. Given the fact that the block grant increases were declining at 1.35% per full-time equivalent per annum and that higher education inflation is higher than the Consumer Price Index, student fees increased at much higher levels than inflation.

There is certainly a need for a study into high tertiary inflation. Two contributors that immediately come to mind are the weakening Rand (import of books and equipment), and inflated salary packages of the ever-increasing cadre of university leadership above professorial level.

With regard to incentives, the undergraduate subsidy system pays universities 70% of the block grant subsidy for enrolments, as well as for institutional factors such as enrolments of disadvantaged students and size of institution, and 16% for graduation completion (the rest is for research and postgraduate outputs).[15] In many countries there is now a debate about shifting the balance between input and output, with some countries discussing a 50-50 split. The low reward for graduation means that universities can take high-risk students, collect 70% of the subsidy and, by inflating fees, cover the cost of the inefficiency of low completion rates. What appears to be a survival strategy (a trade-off between demand for transformation and quality) is not only morally questionable, but also a lose-lose situation for the poor students and the economy.

For the poor students who do not graduate and do not pay back, the National Student Financial Aid Scheme is an extension of the social grant system, but could also be the 'revolving door' outcome against which the White Paper warned in 1997,[16] where poor students are enabled to enter the higher education system, but being unable to complete their studies, are 'revolved' back into poverty, but in this case, with the additional burden of a student loan debt they are unable to repay because they lack the qualifications to secure formal employment. So, rather than higher education being an empowering mechanism, it instead disempowers poor students and puts them deeper into debt. Are we surprised that some of these students went beyond a protest march?

In 2015, both rich and poor students revolted and there is considerable anecdotal evidence that the ones who tried to burn down university administration buildings containing fee records were the ones with bad debt and bad academic records. The students had finally realised that this pretence by government and the higher education system to redress inequality through higher education was not working, and will not work. After all, even Piketty[10] says that higher education does not solve inequalities; it can only keep them from becoming unsettling.

The system must change

For Piketty[10], the best way to reduce inequality and increase the overall growth of the economy is to invest in higher education. He argues that not even minimum wage schedules can multiply wages by, say, factors of five or ten. To achieve that level of progress, education and technology are the decisive factors.

Partrinos[17], from his study of 140 countries, makes three important policy points. Firstly, higher education returns are high and need to be funded better. Secondly, globally, and presumably even more so in South Africa, the high returns will fuel a demand for tertiary education and governments will need to seriously consider appropriate policies for financing this demand. Thirdly, in an environment of high returns to university education, any lowering of private costs means that the general taxpayer (who earns an average income) effectively pays for the education of the rich (who earn an above average income). This confirms the findings from a prominent South African economist that free higher education for all is a policy idea that will harm, rather than assist, the poor.[18]

So then what about free higher education for the poor? The South African government's own report makes a strong case for free higher education for the poor[19] and this should be supported. But the really tough questions are: how will free education be undertaken and for how many? In most countries in the world, developed or developing, a very small proportion of the poor go to university, and ultimately complete successfully, because of lack of academic, social and material capital. There is no evidence anywhere in the world that large numbers of the poor can, through higher education alone, take one giant step into the middle class. China has proportionally invested in higher education at a rate never before observed,[12] but it is not free higher education and the university sector is an integral part of the state's development plan, while

in South Africa, higher education with its high private returns, is clearly seen as individual mobility.

To provide greater access and chances of success to poor students will force South Africa to confront the long-avoided differentiation choices. The first is that in order to maintain the best postgraduate system in Africa and to allow for successful access, universities must be differentiated into institutional types, somewhat like in California, which has the most successful higher education system in the world. In California, there are a range of institutions – from community colleges (remedial schools with some vocational offerings) and undergraduate universities (e.g. Los Angeles South West College) to world-class research universities (e.g. Berkeley and Stanford). This system is also under threat from low taxes and poor financial management.[20]

The key for such a system is strong articulation – something South Africa has talked about for 20 years but has done very little about. Barack Obama started at Occidental College in Los Angeles, transferred to Columbia and then to Harvard. Obama's latest legacy programme is free community colleges. According to a brief issued by The White House, Obama's rationale is:

> *In the coming years, jobs requiring at least an associate degree are projected to grow twice as fast as jobs requiring no college experience. We will not fill those jobs – or keep those jobs on our shores – without the training offered by community colleges.*[21]

In the South African context, this would require a radical rethink of our current notion of a community college, never mind a technical and vocational education and training college.

An alternative is to change the current colonial legacy of a 3-year degree with an honours degree to a 4-year system, with the possibility of a diploma or associate degree exit after two years. The key issue is that the students, as Van den Berg shows, have to leave university with a qualification. Currently, South Africa has a 'have or have not' structure, meaning high returns for degrees or unemployment. The honours degree is a major stumbling block – particularly for black students – because there is limited postgraduate funding for the honours qualification.[22]

If such a model is applied to all universities, the South African higher education system could become a kind of hybrid college/university system. Admittedly, this could have unanticipated consequences, but for a start it would serve the development-equity imperative better than the current system. Perhaps more important than decolonising the curriculum would be restructuring the undergraduate tertiary landscape.

With the highest private returns to higher education in the world, free higher education for all would not only be scandalous, it would destroy the best postgraduate university system in Africa. Higher education should resist the South African Airways bailout approach to fees. The debate should not just be about different models of direct or deferred fee payments, instead the structure of the undergraduate system needs to be rethought within a framework of empirical evidence about the features, and contradictory demands, of the system.

Acknowledgements

I would like to thank Ian Bunting, Charles Sheppard, Daya Reddy and Servaas van den Berg for making inputs to this paper.

References

1. Bothwell E. BRICS and emerging economies rankings 2016 results announced. Times Higher Education World University Rankings. 2015 Dec 02. Available from: https://www.timeshighereducation.com/news/brics-emerging-economies-rankings-2016-results-announced
2. Cloete N, Mouton JM, Sheppard C. The doctorate in South Africa: Discourse, data and policies. Cape Town: African Minds; 2015.
3. Mangcu X. No transformation in higher education institutions. SABC News. 2014 Aug 25. Available from: http://www.sabc.co.za/news/a/db684900453a10f9a7f5b7a5ad025b24/No-transformation-in-higher-education-institutions:-Mangcu-20140825
4. Govinder KS, Zondo NP, Makgoba MW. A new look at demographic transformation for universities in South Africa. S Afr J Sci. 2013;109(11/12), Art. #2013-0163, 11 pages. http://dx.doi.org/10.1590/sajs.2013/20130163
5. Cooper D, Subotzky G. The skewed revolution. Trends in South African higher education: 1988–1998. Cape Town: Education Policy Unit, University of the Western Cape; 2001.
6. Department of Higher Education and Training (DHET). Are we making progress with systemic structural transformation resourcing access, success, staffing and researching in higher education? Paper prepared for the Higher Education Transformation Summit Pretoria: DHET; 2015.
7. Department of Higher Education and Training (DHET). Higher Education Management Information System. Pretoria: DHET; 2013.
8. Council on Higher Education. A proposal for undergraduate curriculum reform in South Africa: The case for a flexible curriculum structure. Pretoria: Council on Higher Education; 2013.
9. Van den Berg S. Inequality, poverty and prospects for redistribution. Dev South Afr. 2015;31(2):197–218. http://dx.doi.org/10.1080/0376835X.2013.871196
10. Piketty T. Capital in the 21st century. Cambridge, MA: Harvard University Press; 2014.
11. Ask My Professor. Piketty on higher education [blog on the Internet]. c2015 [cited 2016 Feb 25]. Available from: http://askmyprofessor.org/piketty-on-higher-education/
12. Montenegro CE, Patrinos HA. Human development reports comparable estimates of returns to schooling around the world. Washington, DC: The World Bank; 2014. http://dx.doi.org/10.1596/1813-9450-7020
13. Harmes L. South Africa's Gini coefficient: Causes, consequences and possible responses. Pretoria: Gordon Institute of Business Sciences; 2013. Available from: http://repository.up.ac.za/bitstream/handle/2263/40181/Harmse_South_2013.pdf?sequence=1
14. GroundUp. Student fees: Facts, figures and observations. Fin24. 2015 Oct 22. Available from: http://www.fin24.com/Economy/Student-fees-facts-figures-and-observations-20151022
15. Ministry of Higher Education and Training. Ministerial statement on university funding 2015/16 and 2016/17. Pretoria: Ministry of Higher Education and Training; 2014.
16. Department of Education. White Paper on Higher Education. Pretoria: Department of Education; 1997.
17. Patrinos HA. Higher education: Returns are high but we need to fund it better. Education for global development [blog on the Internet]. c2015 [cited 2016 Feb 25]. Available from: http://blogs.worldbank.org/education/higher-education-returns-are-high-we-need-fund-it-better
18. Archer S. Free higher education is an inequality engine. Business Day. 2015 Oct.
19. Department of Higher Education and Training (DHET). Report of the working group on fee free university education for the poor in South Africa. Pretoria: DHET; 2012. Available from: https://d3n8a8pro7vhmx.cloudfront.net/amandla/pages/74/attachments/original/1446105100/Final_Draft_Report_of_the_Working_Group_on_Fee-Free.pdf?1446105100
20. The Economist. California's universities in trouble: Before the fall. Economist. 2009 Aug 06. Available from: http://www.economist.com/node/14183037
21. The White House. Building American skills through community colleges [article on the Internet]. no date [cited 2016 Feb 25]. Available from: https://www.whitehouse.gov/sites/default/files/100326-community-college-fact-sheet.pdf
22. Mouton J, Van Lill M, Botha J, Boshoff N, Valentine A, Cloete N, et al. A study on the retention, completion and progression rates of South African postgraduate students. Stellenbosch: Centre for Research on Evaluation, Science and Technology; 2015.

Producing the next generation of water resource experts in South Africa

AUTHORS:
Falko T. Buschke[1]
Marinda F. Avenant[1]
Surina Esterhuyse[1]
Marthie E. Kemp[1]
Fred J. Kruger[1]
Olusola O. Ololade[1]
Maitland T. Seaman[1]
A. Tascha Vos[1]
Marie Watson[1]

AFFILIATION:
[1]Centre for Environmental Management, University of the Free State, Bloemfontein, South Africa

CORRESPONDENCE TO:
Falko Buschke

EMAIL:
falko.buschke@gmail.com

POSTAL ADDRESS:
Centre for Environmental Management, University of the Free State, PO Box 339, Bloemfontein 9300, South Africa

KEYWORDS:
specialists; institutional capacity; historical context; resilience; mindset shift

HOW TO CITE:
Buschke FT, Avenant MF, Esterhuyse S, Kemp ME, Kruger FJ, Ololade OO, et al. Producing the next generation of water resource experts in South Africa. S Afr J Sci. 2016;112(3/4), Art. #a0145, 4 pages. http://dx.doi.org/10.17159/sajs.2016/a0145

Introduction

The 2nd African Water Symposium, in conjunction with the 6th Orange River Basin Symposium, was held on the campus of the University of the Free State, Bloemfontein, on 7 and 8 October 2015. The theme of the symposium was 'systems thinking in environmental water resources management', with the aim of evaluating and debating a holistic approach to water management in southern Africa. The symposium included a panel discussion that was originally intended to define a roadmap towards effective water resources management, but it soon took on a life of its own and evolved into a discussion on how to develop and train the next generation of managers and scientists to tackle the current challenges being faced by the wider water sector. This commentary serves as a record of the main themes identified during the panel discussion and begins by setting the South African context and identifying the plurality of values and opinions held by stakeholders in the post-apartheid water sector. It then elaborates on the decline of water specialists in the country and identifies the characteristics needed from a new generation of water scientists and managers. Finally, it outlines some of the proposed strategies to cultivating a new cohort of specialists able to tackle the challenge of effective water resources management in South Africa.

Shaking off the shackles of history

The 2-hour long discussion session consisted of opening statements from a panel of scientists and practitioners specifically selected for their experience in different parts of the water sector (Table 1). After statements from these experts, the floor was opened to contributions from the audience, which included government officials, university researchers and professionals from environmental non-governmental organisations.

The panel of experts began by highlighting what they felt were the main constraints to effective water management, which included familiar culprits: poor governance caused by absent or ineffective communication between different governmental spheres and departments, lack of deployment and enforcement of our progressive national water policies, stop-start engagement of the new statutory institutions (such as water user associations) and the rapidly deteriorating infrastructure for water provisioning and sewage treatment. Overarching these dilemmas was the looming shortcoming of ever-eroding expertise and institutional capacity. Identifying these problems was certainly not unique; in fact, these same weaknesses were identified during the same symposium four years earlier.[1] However, reiterating these shortcomings during the panel discussion led to an important moment of reflection after the floor was opened to comments from the audience.

One of the major obstacles to effective water resources management became immediately clear after only a few moments of discussion. The first audience member, noting that the panel of experts was made up exclusively of white males at or approaching retirement age, pointed out that their criticisms were perhaps unfounded. Although she acknowledged that the efficacy of national water management has declined during recent decades, she argued that this was not necessarily the result of deteriorating institutional infrastructure. Instead, she reasoned that the water provisioning landscape in post-apartheid South Africa has stretched the water sector to breaking point. *The National Constitution (Act 108 of 1996)* and the *Water Act (Act 36 of 1998)* meant that the government's priority necessarily lay in correcting the historical neglect in servicing the very large number of South Africans with a right to clean drinking water. Inevitably, the unprecedented demand for water quickly outpaced the ability of the state to supply it. Immediately following the democratic elections of 1994, the first post-apartheid government faced a situation where one third of the population (estimated 12 million out of a total 36 million people) did not have access to clean drinking water.[2] A decade later, the government had reduced this number to an estimated 3.7 million people out of a total population of 48.1 million.[2] It was inevitable that the whole water sector would struggle to cope in a water-scarce country that almost doubled the number of citizens with access to clean water while also sustaining a 3% increase in gross domestic product (GDP) and a 1% growth rate in GDP per capita during the same period.[3]

This exchange lead to an instance of conflict in the meeting where the current decline of effective water management was brought face-to-face with the pressures to redress historical injustices. In this, the panel discussion in a small portion of central South Africa became a microcosm for the water sector at large. Fortunately, after a few anxious objections, the tension in the room dissipated after communal *mea culpa*. Those in attendance realised that both the panel of experts as well as the critical commentator from the audience had valid arguments. Just as it is unfair to ignore the historical context in South Africa, is it equally unfair to continue blaming our apartheid past for all our present shortcomings in the water sector.

The lesson learnt, therefore, was that different stakeholders hold different interpretations of the same situation. Moreover, the only way to prevent dialogue from disintegrating into accusations and blame-shifting is to acknowledge the fact that the water sector is a tangled knot of, sometimes contradictory, impressions and perceptions. Unless different groups in South Africa experience, what Brown[4] terms, a 'cultural revolution', then the legacy of apartheid – and the maligned perceptions of and by different groups – will continue to interfere with participatory approaches to water management.

Lamenting the lost limnologists

The panel discussion moved on to more technical topics once those in attendance had acknowledged and accepted the varying opinions of their colleagues in the room. However, the theme of the first comment from the audience lingered on. The experts on the panel were indeed hovering near retirement age. While it is convenient to dismiss this as poor planning by the symposium organisers, the truth is that this lack of diversity is symptomatic of the wider water sector.

Table 1: The panel of scientists and practitioners specifically selected for their experience in different parts of the water sector.

Expert	Expertise	Experience
Dr Harry Biggs	Adaptive management of aquatic ecosystems	Former Programme Manager of the Kruger National Park Rivers Research Programme; Programme Integrator for Systems Ecology for South African National Parks; IUCN/WCPA freshwater task force coordinator
Mr Nic Knoetze	Water usage in the agricultural sector	Former Deputy Regional Director for the Department of Water Affairs and Forestry in the Northern Cape Province. Currently Chief Executive Officer for the South African Association for Water User Associations (SAAFWUA)
Prof. Maitland Seaman	Aquatic ecology, biomonitoring training and education	Former researcher at the National Institute for Water Research and recently retired Director of the Centre for Environmental Management at the University of the Free State
Prof. Anthony Turton	Water governance and strategic planning.	Formerly at the national Council for Scientific and Industrial Research (CSIR) in the position of Strategic Research Leadership: Water Resources Management; Currently Professor in Environmental Management at the University of the Free State
Dr Johan van der Merwe	Geohydrology and water provisioning	Former Deputy Director at the Department of Water Affairs and Forestry (Water Quality and Geohydrology) in the Free State and Deputy Director of Strategic Support in the Department of Water Affairs. Currently an Affiliated Researcher at the Institute for Groundwater Studies at the University of the Free State

Harding[5] reported that from as early as 1989, scientists have complained about the insufficient financial support for South African limnologists, which at the time was less than that for a single Australian institute. He further argued that this lack of investment has not changed over the past 25 years, leading to many potential limnology students gravitating towards more richly-funded fields like chemistry and microbiology. This has resulted in a lost generation of limnologists with the necessary skills to grapple with water-related issues.

The heading of this section is restrictive for alliterative purposes, but it should not imply that the shortage of expertise is limited to limnology. South Africa also has a limited number of hydrologists and geohydrologists. All of these water specialisations are necessary to develop and manage South Africa's limited water resources efficiently. With more than 80% of our surface water resources already allocated to water users,[6] the need to develop and manage groundwater resources (of which approximately 3 500 Mm^3/a is available for further development) is critical. This seems the most promising way to address water deficits, which have already been reported in half of the water management areas in South Africa.[7] The challenge of meeting the current and future water demands, while redressing injustices of the past, can only be achieved by developing groundwater resources in tandem with managing surface water resources more efficiently. Moreover, surface and groundwater should be managed as one holistic system. To achieve this goal, we need to equip limnologists, hydrologists and geohydrologists with a skillset to develop and manage the water resources in the face of environmental and socio-economic change.

The sad reality, however, is that the number of freshwater scientists with active research programmes has declined in South Africa since the mid-1990s.[8] Furthermore, even when universities still offer study programmes in limnology, hydrology and geohydrology (as is the case at the University of the Free State), these programmes are offered independently and are spread across multiple departments. This thwarts any attempts to integrate surface and groundwater into one holistic system.

During a recent interview, the Minister of Water and Sanitation, Nomvula Mokonyane, stated that 'we need more bodies with the knowledge and capability...the right people in the right place with the ability to do the job'.[9] Yet, despite this encouraging endorsement, the estimated percentages of vacant posts in the Department of Water and Sanitation for hydrogeologists and geohydrologist in 2010 was 47% and 53%, respectively.[6] One reason for this is the difficulty of replacing recently retired experts, so much so that in 2010 more than 50% of the groundwater personnel at the Department of Water and Sanitation had fewer than 5 years' experience and lacked experienced mentors to guide them.[6]

The dearth of specially-trained water scientists was echoed during the panel discussion with many of the subsequent comments criticising the work-preparedness of recent graduates who, it was argued, lack the necessary deep technical understanding of the complexities of the water sector. Moreover, it was suggested that graduates are unable to respond quickly enough to the ever-changing demands of the water sector. This, therefore, begs the question: is it possible for individual students to embody a deep understanding of complex topics, have the ability to adapt to changing circumstances as well as the deliberative skills for coping with the diverse values and perceptions of their colleagues and stakeholders?

There are currently two predominant paths toward employment in the water sector. The first is the traditional path through a single academic discipline, such as civil engineering or aquatic ecology. The second path is one through modern transdisciplinary programmes, such natural resources management, public policy or environmental management. The trouble is that the former group often lacks the over-arching perspective of the broader sector whereas the latter group is missing the deep expertise to solve complex technical issues. Consequently, the path forward for training water professionals needs to amalgamate both core disciplinary competencies and a holistic appreciation of the broader sector together with the competence required to address complexity, all within a single curriculum.

Nurturing new water resource expertise

Being an academic symposium, the panel of experts as well as many members of the audience had their foundations in a single academic discipline. However, as the discussion progressed it became clear that few of these people were constricted by their academic background. Amongst the expert panellists was, for example, Dr Harry Biggs who began his career as a veterinarian before moving into adaptive

management in complex systems and Prof. Anthony Turton, who started out in the National Intelligence Services before branching out into water-related issues. Furthermore, even though the other panellists largely remained in the discipline in which they were originally trained (engineering, geohydrology and aquatic ecology), their career paths included forays into water management, policy and provisioning. Each of their cases typified 'T-shaped' skills, where a depth of understanding is initially developed within a specialist discipline, but is then coupled with a capability to understand and interact with specialists from a wide range of fields.[10]

How can we encourage the proliferation of people with T-shaped skills in the water sector? Perhaps a suitable analogy – and one befitting the theme of the symposium – relates to how the management of natural systems has moved away from viewing nature as a static phenomenon to viewing it as dynamic and adaptable instead. Consequently, management interventions no longer try to keep these ecosystems in a fixed state, but rather allow for complexity, dynamism and resilience. Resilience in ecosystems is viewed as the capacity of a system to withstand shocks while still retaining it essential function and structure.[11] High-altitude wetlands, non-perennial rivers, man-made dams and estuaries might have completely different functions and structures, but their resilience can be defined by the same set of characteristics: slowly changing state variables, the extent of endogenous self-organisation (as opposed to external drivers) and mechanisms for the evolution of novelty.[12] Similarly, specialists in the water sector should not be pushed into constrictive boxes defined by specialist academic training. Instead, we should encourage these individuals to become resilient and adaptable to changing contexts, without jeopardising their core competencies.

Fazey and colleagues[13] propose that, like socio-ecological resilience, resilient individuals can be classified according to four main requirements. The first requirement is the willingness to maintain resilience, which is distinct from exclusively pursuing productivity and efficiency. This necessitates a conceptual change where innovation and flexibility are granted the same level of importance as efficiency and optimisation.[14] The second requirement is an awareness of current problems as well as the desired endpoint. This is where a deep technical knowledge of a specific discipline interacts with philosophical judgements of what is valued and desired. Moreover, it allows for value pluralism – and regular conflict between values – by acknowledging that utilitarianism and instrumentalism are only two of many ethical positions.[15] The third requirement is proactive behaviour, which borrows from post-normal science.[16] Post-normal science, unlike normal science, which is settled in a fixed paradigm, does not view ignorance as negative or threatening, but rather as an essential complement to knowledge.[16] It also means rethinking future uncertainties and not making misinformed assumptions about the future state of the environment.[15] The final requirement for a resilient individual is the ability to change existing behaviours. This requires the existentialist viewpoint that the narrative of any individual – the way she sees and fulfils her own role within society – is changeable through self-determination.[17]

The atmosphere in the panel discussion suggested that South Africa is failing to produce adaptive and resilient water professionals. This is likely because tertiary education institutions are not teaching students different ways of thinking for a variety of situations (i.e. metacognition).[13] Instead, increasing class sizes (without proportional increases in funding) are forcing universities to rely on less qualified staff using automated assessment methods,[13] which ultimately shifts the focus away from developing metacognition and towards the regurgitation of the facts and knowledge from other people's thinking. One can only imagine that the increasing financial burden on universities in the aftermath of the #feesmustfall protests in the higher education sector will only weaken the capacity of universities to invest in developing resilient and adaptable graduates.

The reality is that we cannot predict in what state the water sector will be when the current generation of students graduate. As the higher education sector, we should, therefore, be preparing our graduates for uncertain futures and the only way to do this is by focusing on producing resilient individuals, not fact-spitting parrots.

The path forward

The South African water sector still faces many challenges; amongst them are the biophysical constraints caused by supplying an increasingly polluted resource to meet an ever-growing demand. Coupled to this is the uniquely South African socio-economic realities typified by the need to redress historical injustices and alleviate wide-scale contemporary poverty. These are undoubtedly 'wicked' problems involving complex and unpredictable systems that contain stakeholders with conflicting interests.[18,19] Solving such problems will require political will and intellectual resolve. Perhaps even more unsettling is the scarcity of well-trained adaptive individuals that can meet the challenges of water resources management in this country. To this end, the onus is on the higher education sector in South Africa to not only expand the training opportunities for aspiring water specialists, but to fundamentally alter the way we train our students. There has to be mindset shift away from producing a mob of superficially-trained graduates to producing a cohort of deep reflective thinkers.

The current water landscape in South Africa is approaching a state of crisis. In this metaphorical war, we need a taskforce of specially-trained operatives, not a troop of cadet soldiers. To reach this goal, emphasis should be placed on drawing in the expertise of the few remaining water experts to share their knowledge and experience by mentoring the upcoming generation. More importantly, higher education institutions should stop fighting over pass rates and how these figures relate to government subsidies. Instead, we should invest our time and energy into nurturing resilient individuals with the capacity to adapt to future uncertainties. Unless we do so, we will face several years of drought in terms of both water availability and intellectual capacity.

Acknowledgements

We would like to thank everyone who attended the panel discussion during the 2nd African Water Symposium their valuable inputs.

References

1. Buschke FT, Esterhuyse S. The perceptions of research values and priorities in water resource management from the 3rd Orange River Basin Symposium. Water SA. 2012;38:249–243. http://dx.doi.org/10.4314/wsa.v38i2.10
2. Muller M. Free basic water – a sustainable instrument for a sustainable future in South Africa. Environ Urban. 2008;20:67–87 http://dx.doi.org/10.1177/0956247808089149
3. Du Plessis S, Smit B. Economic growth in South Africa since 1994. Stellenbosch Economic Working Papers: 1/2006. Stellenbosch: Department of Economics; 2006.
4. Brown J. Assuming too much? Participatory water resource governance in South Africa. Geogr J. 2011;177:171–185. http://dx.doi.org/10.1111/j.1475-4959.2010.00378.x
5. Harding WR. Living with eutrophication in South Africa: A review of the realities and challenges. T Roy Soc S Afr. 2015;70:155–172. http://dx.doi.org/10.1080/0035919X.2015.1014878
6. Department of Water Affairs. Groundwater Strategy 2010. Pretoria: Department of Water Affairs; 2010.
7. Department of Water Affairs. National water resource strategy. 2nd ed: Water for an equitable and sustainable future. Pretoria: Department of Water Affairs; 2013.
8. Ashton PJ, Roux DJ, Breen CM, Day JA, Mitchell SA, Seaman MT, et al. The freshwater science landscape in South Africa, 1900–2010: Overview of research topics, key individuals, institutional change and operating culture. Report No. TT 530/12. Pretoria: Water Research Commission; 2012.
9. Barron C. So many questions. Sunday Times. 2015 November 08; p.21. http://dx.doi.org/10.1016/B978-0-08-100250-6.00007-9
10. Cornell S, Berkhout F, Tuinstra W, Tàbara JD, Jäger J, Chabay I. et al. Opening up knowledge systems for better responses to global environmental change. Environ Sci Policy. 2013;28:60–70. http://dx.doi.org/10.1016/j.envsci.2012.11.008
11. Holling CS. Resilience and stability of ecological systems. Annu Rev Ecol Syst. 1973;4:1–23. http://dx.doi.org/10.1146/annurev.es.04.110173.000245

12. Carpenter S, Walker B, Anderies JM, Abel N. From metaphor to measurement: Resilience of what to what? Ecosystems. 2001;4:765–781 http://dx.doi.org/10.1007/s10021-001-0045-9

13. Fazey I, Fazey JA, Fischer J, Sherren K, Warren J, Noss RF, Dovers SR. Adaptive capacity and learning to learn as leverage for social-ecological resilience. Front Ecol Environ. 2007;5:375–380. http://dx.doi.org/10.1890/1540-9295(2007)5[375:ACALTL]2.0.CO;2

14. Fazey,I. Resilience and higher order thinking. Ecol Soc. 2010;15(3):9.

15. Spash CL. The shallow or the deep ecological economics movement? Ecol Econ. 2013;93:351–362. http://dx.doi.org/10.1016/j.ecolecon.2013.05.016

16. Funtowicz S, Ravetz J. The emergence of post-normal science. In: Von Schomberg R, editor. Science, politics and morality. Scientific uncertainty and decision making. Dordrecht: Springer Science and Business Media; 1993. p. 85–126. http://dx.doi.org/10.1007/978-94-015-8143-1_6

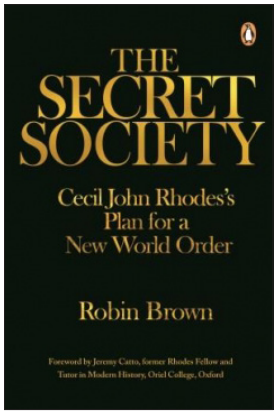
17. Child MF. Conservation of adaptive self-construction: A flux-centred solution to the paradox of nature conservation. Environ Val. 2011;20:527–548. http://dx.doi.org/10.3197/096327111X13150367351339

18. Churchman CW. Wicked problems. Manag Sci. 1967;14:B141–B142.

19. Camillus JC. Strategy as a wicked problem. Harvard Bus Rev. 2008;86:98–106.

Secret society, secret sources?

BOOK TITLE:
The secret society: Cecil John Rhodes's plan for a new world order



AUTHOR:
Robin Brown

ISBN:
9781770229204 (hardcover)

PUBLISHER:
Penguin, Cape Town; ZAR350

PUBLISHED:
2015

REVIEWER:
Iain Edwards

EMAIL:
edwards.researchsa@gmail.com

AFFILIATION:
Gordon Institute of Business Science, University of Pretoria, Pretoria, South Africa

POSTAL ADDRESS:
GIBS, PO Box 787602, Sandton 2146, South Africa

HOW TO CITE:
Edwards I. Secret society, secret sources? S Afr J Sci. 2016;112(3/4), Art. #a0150, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0150

Robin Brown's new book concerns Cecil John Rhode's, southern Africa's greatest jingo and racial supremacist, and a figure whose afterlife continues to bedevil South African's attempts to come to terms with their past. Published but months after the #RhodesMustFall movement began, Brown's South African book launch came with rousing prepublication and publication media fanfare with excerpts of the book appearing across the range of South African media.

Brown's book is not simply another in a long stream of biographies of Rhodes, which although differing in political perspective and historical judgement, all follow a chronological narrative and are remarkably similar in content. Claiming to draw upon hitherto hidden or recently discovered archival sources, Brown's work is episodic, not chronological. The book is as much about Rhodes, who died in 1902, as it is about his afterlife. Included are Rhode's ideas for a secret society to ensure that English-speaking societies worldwide fashioned and controlled a global New World Order.

Brown emphasises three issues: one is already known, one long insinuated, and the final one is Brown's *pièce de résistance* which form the thread through *Secret Society*: Rhode's wills and attached testimonies, Rhode's homosexuality; and, posthumously, a secret society of powerful and mostly homosexual men wielding unheralded behind-the-scenes power and influence in British and global politics. For Brown, Rhodes is seeking to shape the future – making history – from beyond the grave.

In a dramatic statement on Rhode's intentions, Brown quotes Rhodes as writing in his Confession, in 1877:

> *I leave all my worldly goods in trust (to the Secretary for the Colonies, Lord Carnarvon, and Sydney Godolphin Shippard now of the Inner Temple) to and for the establishment, promotion and development of a Secret Society, the true aim and object thereof shall be the extension of British rule through the world, the perfecting of a system of immigration from the United Kingdom, and of colonisation by British subjects of all lands where the means of livelihood are attainable by energy, labour and enterprise, and especially the occupation by British settlers of the entire Continent of Africa, the Holy Land, the Valley of the Euphrates, the islands of Cyprus and Candia, the whole of South America, the Islands of the Pacific not heretofore possessed by Great Britain, the whole of the Malay Archipelago, the seaboard of China and Japan, the ultimate recovery of the United States of America as an integral part of the British Empire, the inauguration of a system of Colonial representation in the Imperial Parliament which may tend to wield together the disjointed members of the Empire, and, finally, the foundation of so great a Power as to hereafter render wars impossible and promote the best interests of humanity.* (p. 42)

Brown refers to this as 'deranged' (p. 42). In South African press releases at the time of publication, this excerpt featured prominently. Brown provides no citations for this quote and concedes that 'it is unlikely to be entirely [Rhode's] own work' (p. 42). Whatever the questionable provenance is, this excerpt is not from Rhodes Confession of 1877, a full copy of the 'fair copy' version of which is produced in an appendix in John Flint's 1975 biography of Rhodes.[1]

Key to this secret society is the Round Table Movement, and Milner, second chairperson of the Rhodes Trust. It was this trust, and Bailey's 'millions', which funded the formation of the Royal Institute of International Affairs at Chatham House in 1920. It was from here that, according to Brown, the Round Table exerted much influence over the ending of the Great War (p. 303–309), the formation of the League of Nations (p. 316), and the Commonwealth (p. 318) and the editorial policy of *The Times*. In addition, the Round Table was of key influence in British policy towards Palestine, India and Ireland, as well as in issues relating to the British appeasement policy, the Windsors and abdication, Hitler's foreign policy, and attempts to retain the empire after 1945.

When dealing with a secret society and the Round Table, Brown assumes the existence and importance of this group, then puts evidence in place to confirm such a priori assumptions. Brown continually uses words such as 'perhaps' and 'almost certainly' (p. 185), 'more than likely' (p. 204), 'highly likely', 'unlikely', 'may even', and 'only logical scenario' (p. 210–211). Scholarly analysis is not constructed by repeated inference and supposition.

Rhodes was fascinated by a secret society run along semi-religious lines. He was initiated into Freemasonry whilst at Oxford, yet, in his Confession of 1877, Rhodes clearly expressed his disdain for Freemasonry, as lacking clear purpose and vision beyond itself. However, to take the existence of a secret society, as initiated by Rhodes, and draw a posthumous *continuum* to Chatham House, the Round Table Movement and with these a hidden hand wielding power of global reach, is another matter.

There are high normative requirements for scholarly studies, most particularly for authors making claims as Brown does. *Secret Society* lacks a clear analytical structure and is hugely overwritten and repetitive. Readers are continually informed, even on page 329 of a 360-odd page text, that Reginald Baliol Brett – Lord Esher – was a royal whisperer and pederast. Brown displays an astonishingly patronising view of black Africans (p. 53, p.60). Chapters have characteristics of hurried drafts and Brown is often factually incorrect. Bill Clinton's presidential nomination speech was in 1992, not 1981. All 28 chapters have footnotes, but the vast majority are chatty asides, not citations. Throughout the book, Brown quotes, often extensively, from sources without providing citations. Chapter 18 is replete with primary source quotes, but has only two footnotes (p.367), neither of which is a citation.

Brown has failed to consult key texts, including recent studies, and is unaware of relevant criss-crossing scholarly traditions. Many of Brown's secondary source citations are carelessly unsystematic. The book has no bibliography, let alone a properly ordered one. Did Penguin not notice these highly serious scholarly lapses?

This is compounded by Brown's attitude to primary source material. Granted, secret societies are by definition not open societies. Evidential and empirical care is essential. Unsubstantiated claims are ethically and politically irresponsible, and can be dangerous, particularly so as conspiracy theorists have long taken it for granted that Rhodes, Milner and others in the Round Table were Freemasons bent on world domination.[2-5]

Brown cites no material from at least three important substantial primary collections. Staggeringly, the first two concern Rhodes directly. The first is Rhodes's voluminous personal papers, long housed in Rhodes House and now part of the Weston Library of the Bodleian Library at Oxford.[6] The second is Rev. John Verschoyle's *Cecil Rhodes: His political life and speeches, 1881–1900,* a collection held in high scholarly regard.[7] Brown mentions Verschoyle, of an ancient aristocratic family, but only as a 'dubious vicar', as one of Rhode's favoured male companions, and as a leading member of the aesthete movement mixing in High Society establishment homosexual circles linked to government and the Secret Society. Yet, Brown provides no footnotes which refer to primary evidence or other sources backing up these claims. The third is the British government's Colonial Office archive, which includes material highly critical of Rhodes's activities.

There may be at least one key source (p. 196–198, p.304). After Lord Esher's death, his son Maurice collected and compiled his father's writings into two 'massive volumes' (p. 196). Brown writes:

> *In Volume One of two rare first editions I was fortunate to acquire, I discovered entries by Regy Brett which record conversations with Rhodes and Chamberlain. These exchanges alter the course not only of imperial history but also, quite possibly, world history* (sic). (p. 196)

Brown simply and inadequately cites this key source as 'Brett's Journals' (Chapter 16, footnote 4). However, as Esher died in 1930, these journals can hardly account for all the unreferenced primary source material Brown uses so extensively to posit for his Secret Society influence in the first half of the 20th century.

In his acceptance speech at the Democratic National Convention in 1992, Clinton made positive mention of Carrol Quigley, one of his professors whilst an undergraduate at Georgetown University. Clinton noted that Quigley 'said to us that America was the greatest country in the history of the world because our people have always believed in two things: that tomorrow can be better than today and that every one of us has a personal, moral responsibility to make it so'.[8] In his autobiography, Clinton expands on why Quigley made such an impression on him, to the extent that '[f]rom the 1992 campaign through my two terms in office, I quoted Professor Quigley's line often, hoping it would spur my fellow Americans, and me, to practise what he preached'. Clinton's reasons for respecting Quigley are perfectly appropriate. Quigley was an eminent and inspiring scholar of civilisations.[9] However, at the time and shortly after the Democratic Convention, Clinton's attitude to Quigley aroused curious comments. For Quigley, who died in 1977, was not only a renowned theorist of the evolution of civilisations, but also intent upon exposing an Anglo-American elite cabal influencing global affairs. Amongst Quigley's other public supporters was the ultra-conservative John Birch Society. To many, Quigley was not only an historian, but also a conspiracy theorist.[10]

Professor Quigley's main area of research was on the Rhodes–Milner Round Table group and its influence on British, Anglo-American and global affairs in the first half of the 20th century. Quigley's research was first published, seemingly obscurely, in 1949, but then later published in 1981.[11] Brown acknowledges Quigley's work, citing it twice (p. 237–238, p.260). Quotes from unacknowledged primary sources appearing in Quigley's work also appear, unacknowledged, in Brown's work. (e.g. p. 300, p. 317). Quigley did not write on homosexuality. He did have a secret source within The Round Table Movement, but never uses the term the New World Order.

Robin Brown needs to provide a professionally listed bibliography of consulted sources; clarify, in accordance with scholarly protocols, the provenance of his references and quotes; and explain why core primary and secondary sources – surely material essential to his dramatic publicity-claiming assertions – show no signs of having been consulted. Failing this, *Secret Society* has no legitimacy.

References

1. Flint J. Cecil Rhodes. Boston, MA: Little, Brown and Company; 1975. p. 248–252
2. The Rhodes ~ Milner Round Table [article on the Internet]. No date [cited 2016 Feb 26]. Available from: https://watch.pair.com/roundtable.html
3. Dubay E. Cecil Rhode's round table groups. The Atlantean Conspiracy [article on the Internet]. c2012 [cited 2016 Feb 26]. Available from: http://www.atlanteanconspiracy.com/2008/06/rhodes-round-table-groups.html
4. Unbanned Bible Publications [article on the Internet]. No date [cited 2016 Feb 26]. Available from: http://www.unbannedbiblepublications.com/index_files/Page2982.htm
5. Illuminati Conspiracy Archive. Who controls the world? The round table group: Googling the elite's who's who [article on the Internet]. No date [cited 2016 Feb 25]. Available from: http://www.conspiracyarchive.com/Commentary/Elite.htm
6. Bodlean Library. Correspondence of Cecil John Rhodes: 1875-1908 [collection on the Internet]. c2011 [cited 2016 Feb 26]. Available from: http://www.bodley.ox.ac.uk/dept/scwmss/wmss/online/blcas/rhodes-cj1.html;
7. Verschoyle F. Cecil Rhodes; his political life and speeches, 1881–1900. London: Chapman Hall; 1900.
8. Clinton B. My life. London: Hutchinson; 2004, p. 78. http://dx.doi.org/10.1136/jmh.2004.000155
9. Quigley C. The evolution of society. An introduction to historical analysis 2nd Ed. Indianapolis, IN: Liberty Fund Incorporated; 1979.
10. Sailer S. Carroll Quigley's conspiracy theory: The Milner group. The Unz Review: An alternative media selection [article on the Internet]. c2015 [cited 2016 Feb 26]. Available from: http://www.unz.com/isteve/carroll-quigleys-conspiracy-theory-the-milner-group/
11. Quigley C. The Anglo-American establishment. New York: Books in Focus; 1981.

Complexities and contradictions of doctoral education in South Africa

BOOK TITLE:
Doctoral education in South Africa

AUTHORS:
Nico Cloete, Johann Mouton and Charles Sheppard

ISBN:
9781928331001 (softcover)

PUBLISHER:
African Minds, Cape Town; ZAR220, (eBook is freely available)

PUBLISHED:
2015

REVIEWER:
Merridy Wilson-Strydom

EMAIL:
wilsonstrydommg@ufs.ac.za

AFFILIATION:
Centre for Research on Higher Education and Development, University of the Free State, Bloemfontein, South Africa

POSTAL ADDRESS:
Centre for Research on Higher Education and Development, University of the Free State, PO Box 339, Bloemfontein 9300, South Africa

HOW TO CITE:
Wilson-Strydom M. Complexities and contradictions of doctoral education in South Africa. S Afr J Sci. 2016;112(3/4), Art. #a0147, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0147

I read this lucidly written and empirically rich book with much interest as someone who both researches higher education and grapples with many of the supervision challenges discussed. The experienced authors of *Doctoral Education in South Afric*a have provided a valuable overview of the central issues and topics relevant to doctoral education internationally and then applied more specifically to the African and South African contexts. The central thesis of the book is carefully mapped out in the first chapter. The thesis is that 'Four imperatives intersect in current debates on the production of PhDs in South Africa. These four discourses concern global and national competition (the imperative for growth), efficiency, transformation and equality' (p.20). The analytical framework that is constructed on the basis of these four imperatives 'capture[s] the ecology (the external demand and accountability environment) of doctoral education and training in South Africa today'(p.23). Particularly useful is the manner in which the framework takes account of the external factors as well as the dynamics between students and supervisors within universities. These four imperatives, or discourses, are then used as the basis for organising the empirical chapters of the book.

Firmly grounded in detailed and rigorous empirical analyses – both quantitative and qualitative – and drawing on several studies on doctoral education conducted by the Centre for Higher Education Trust and the Centre for Research on Evaluation, Science and Technology over several years, the authors draw powerful (and likely somewhat controversial) conclusions regarding what it will take to reach the ambitious National Development Plan[1] target of 100 doctoral graduates per million of the population by 2030. Amongst others, the need to tackle the thorny issue of differentiation in the system is discussed, as is the tension between the significant growth of PhD enrolments (6.4% from 1996 to 2012) compared to only 2.9% growth for academic staff in the same period. Chapter 6, which presents an analysis of specific university departments that have been particularly productive with respect to doctoral education, shows the various ways in which the actors in the system have been responding to the contradictions of increasing numbers of PhDs without the requisite increase in resources. The examples and practices shared are of particular value to doctoral supervisors and departments who are seeking to improve their own practices. The major challenge created by the fact that more than 60% of South African PhD students study part-time is also highlighted as one of the main contributors to low completion rates, and the data show the huge impact this situation has on the entire doctoral education system. This finding raises critical issues related to system efficiency and, as is argued in the book, requires that the models of PhD education in the country be revisited.

What might these revised models look like? The authors argue that the system could either opt for incremental change, which would involve continuing with the practices that have shown results in the high performing departments, or could embrace a more radical approach that requires changing the dominant model of doctoral education in South Africa. The latter is presented as the preferred option. This new model (also called a paradigm shift in the book) would involve reversing the full-time to part-time student ratio such that 60% of students would be full-time students. Suggestions for how this might be done are presented, and the realities of the funding requirements (estimated to be about ZAR800 million per year) for such a shift are briefly addressed. The argument would have been strengthened though if more attention was given to possible funding models, particularly given the major resource constraints faced at all the levels in the post-school sector. The final chapter presents three policy options that emerge from the earlier chapters. The first centres on growing doctoral enrolments and graduates (including the setting out of 16 theoretical scenarios). The second policy option is a proposal to make South Africa a PhD hub for Africa, and the third policy is to implement more active differentiation of the sector, allowing for targeted investment in doctoral education, the provision of which is already differentiated across different groupings of universities. None of the options presents a panacea, and each raises a series of tough questions for policymakers and universities. As is emphasised in the concluding section of the book, the research has highlighted a need for better consideration and management of the policy trade-offs of any given policy position. These trade-offs are clearly articulated in the sections mapping out the three main policy options presented.

What is less clear from the arguments presented throughout the book is the normative position that a country like South Africa ought to take up in making these difficult policy trade-off decisions. Although there is some acknowledgement of wider purposes for the doctorate (raised particularly by some of the commentators included in Appendix 2), it seems that the main purpose of the doctorate is framed as contributing to building the knowledge economy and to economic development in South Africa, and Africa. However, what of the complex debates in the higher education landscape about the public good role of the university that have so powerfully been raised by students in the past few months?

Although perhaps beyond the scope of this book, but nonetheless critical when thinking about doctoral education in South Africa, and particularly what this means in terms of larger questions about the purpose of higher education in the country, we ought to also ask pressing questions about knowledge itself: 'What knowledge is produced? How has it been produced? Whose interests does it serve? And how does it serve society?'[2] These questions are particularly important to answer given the current juncture in the country's history, at which young people are increasingly standing up to pose critical questions about the colonial histories and persistent legacies within our universities as a basis for advocating for deeper change within the sector. In my reading of the analysis and arguments presented here, a possible gap is that little consideration has been given to the role that doctoral education has or could play in either subverting or maintaining the colonial heritage of our universities, and its role

in addressing broader social development imperatives of the country. Related points are made in Appendix 2 by Badat, Moja and Langa. Nonetheless, this important and well-researched book certainly takes the debate forward in meaningful ways, and clearly sets out the policy implications of different paths that might be considered as we continue to strive to improve doctoral education in South Africa. The data, conclusions, recommendations, and additional information included in the detailed appendices, are likely to be of much value across the sector, for doctoral students, supervisors, university management and leaders, and policymakers.

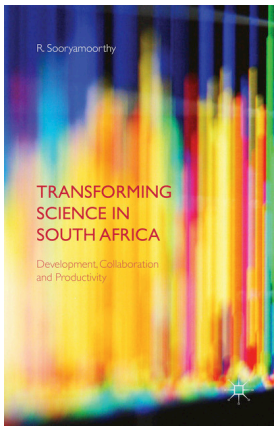
References

1 National Planning Commission (NPC). National Development Plan 2030: Our future, make it work. Executive Summary. Pretoria: NPC; 2012. Available from: http://www.gov.za/sites/www.gov.za/files/Executive%20Summary-NDP%202030%20-%20Our%20future%20-%20make%20it%20work.pdf

2 Herman C. Doctoral education in South Africa – Research and policy. Perspect Educ. 2011;39(3):i–v.

'No man is an island, entire of itself ...'

BOOK TITLE:
Transforming science in South Africa: Development, collaboration and productivity



AUTHOR:
Radhamany Sooryamoorthy

ISBN:
9781137493064 (hardcover)

PUBLISHER:
Palgrave Macmillan, London; EUR89.99

PUBLISHED:
2015

REVIEWER:
Anna Crowe

EMAIL:
aacrowe@hotmail.com

AFFILIATION:
Former lecturer in the Department of Zoology and the School of Education, University of Cape Town, Cape Town, South Africa

POSTAL ADDRESS:
8 Alster Avenue, Newlands 7700, Cape Town, South Africa

HOW TO CITE:
Crowe A. 'No man is an island, entire of itself ...' S Afr J Sci. 2016;112(3/4), Art. #a0149, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0149

Science and scientists can play vital roles in society. Both have contributed in varying degrees to the success of nations, and the frontiers and multidimensionality of science are being extended at rates unprecedented in history. Therefore, although opportunities abound, each new development makes it more difficult for individual scientists to conduct cutting-edge research on their own. This opens, and indeed necessitates, opportunities for scientific collaboration, if the advancements of science are to be used fully to society's advantage. In short, the era of scientists acting as intellectual islands is over.

I anticipated reading *Transforming Science in South Africa: Development, Collaboration and Productivity,* hoping that I would find reasoned explanations about how South African scientists might adapt their approaches to science collaboratively, so that they can effect such transformation. In particular, I was excited to know more about 'the South African model' of scientific collaboration advertised on the back cover of the book. As the author of this book, Professor Sooryamoorthy, is a sociologist, I was also interested in viewing the practise of science from a sociological perspective.

The book includes Sooryamoorthy's research about scientific collaboration and the research productivity of scientists, some of which has been previously published as academic papers. The historical review of science in Africa and in South Africa in particular emphasises changes in levels of collaboration between scientists and identifies a body of literature potentially of value to scholars interested in the general development of science in Africa.

One chapter is devoted to developing a conceptual understanding of scientific collaboration, but there is no explicit explanation as to how this concept is specifically applicable to the remaining chapters of the book. An explicit conceptual or theoretical framework which connects the nine chapters together would have made the book a coherent story.

A number of different types of research methodologies are used to explore relationships between the scientific collaboration and productivity of South African scientists. A bibliometric study of the research publications of South African scientists from 1945 to 2010 identifies key characteristics relevant to scientific collaboration and productivity. These key characteristics are used to further explore qualitative and quantitative relationships between scientific collaboration and productivity the period 1975–2010. In-depth interviews with 204 scientists from one province, KwaZulu-Natal, are analysed qualitatively and quantitatively to determine relationships between the previously identified key characteristics of scientific collaboration and productivity. The findings of these analyses are supported by a qualitative analysis of an in-depth interview with one eminent scientist, also from KwaZulu-Natal.

However, the statistical techniques described and the statistical results produced are confusing in how they are presented and are therefore difficult to interpret in the way that the author has interpreted them. The extremely low r^2-values (coefficient of determination – i.e. the amount of information explained by the regression) reported in the regression analyses used to assess relationships between aspects of the scientists' collaboration and productivity cannot be used to support some of Sooryamoorthy's conclusions. For example, these analyses do not conclusively support that 'the data permit prediction of the degree of collaboration of scientists in the basis of certain known [identified] factors', or that 'collaboration has shown its strong connection to productivity' in the selected sample of scientists. Further, the regression analyses suggest that other important aspects of scientific collaboration and productivity have yet to be identified.

Although the way that scientists collaborate in South Africa might be 'unique' when compared with how scientists collaborate in the rest of Africa, the author fails to present sufficient evidence or arguments to convince the reader that the model which he describes is sufficiently different from international best practices to warrant being called 'the South African model'.

This book was difficult to read for a number of reasons. The text is dense, detailed and sometimes unnecessarily repetitive within paragraphs and sections of chapters. Often statistics are presented in the text when they would have been easier to assimilate if the information had been summarised in a table. Furthermore, some complex tables are unnecessary because they contain only one value from the table that is referenced (and repeated) in the text. Detailed footnotes are used, but their explanations are only given towards the end of the book. This breaks a reader's concentration. Some sentences begin with 'it' or 'they' and the reader has to pause momentarily while they figure out to what 'it' or 'they' refers. Some paragraphs start with acronyms used as nouns (e.g. 'HBUs ...') – which is jarring, especially when the acronym is not in common use. In some instances, terms and concepts are used without qualifying what they mean. For example, 'Mode 2' is used to describe a new form of knowledge production, but no explanation is given for the meaning of Mode 2. Terms and concepts which are subtly different are used interchangeably – for example, 'relevant' and 'significant'.

More problematic is that sometimes, but not always, Social Sciences, and/or Medical Sciences, and/or Veterinary Sciences and Mathematics are combined in the data analyses or the statistics reported. The branches of Sciences are defined differently within and between chapters. For example, Natural Sciences is often reported separately from Physics and Chemistry within a comparison, despite the fact that Physics and Chemistry are Natural Sciences. Consistent use of a defined set of branches of science would have made comparisons between sets of data possible and valid.

The term 'scientists' is inconsistently used in the book. For example, categorising 'scientists' as either 'academics' or 'scientists' in the quantitative analyses is confusing at best, and dangerously sloppy at worst.

Independently of its intellectual value, the book provides a valuable accumulation of a spectrum of knowledge about how science was, and is, practised internationally and within South Africa. However, the price of this book in rands will probably make it unaffordable to most South Africans.

Curiously, given that collaboration is the central theme of this book, Sooryamoorthy chose to write it on his own. This book would have benefitted had it been written together with a scientist and a statistician.

Title of review taken from: Meditation *XVII by John Donne (1572–1631)*

Treatment technology for brewery wastewater in a water-scarce country: A review

AUTHORS:
Adesoji T. Jaiyeola[1]
Joseph K. Bwapwa[1]

AFFILIATION:
[1]Department of Civil Engineering and Surveying, Mangosuthu University of Technology, Durban, South Africa

CORRESPONDENCE TO:
Adesoji T. Jaiyeola

EMAIL:
soj707@yahoo.com

POSTAL ADDRESS:
Department of Civil Engineering, Mangosuthu University of Technology, PO Box 12363, Jacobs 4026, Durban, South Africa

DATES:
Received: 17 Feb. 2015
Revised: 11 June 2015
Accepted: 18 Aug. 2015

KEYWORDS:
brewery effluent; anaerobic treatment; aerobic treatment; membrane technology; activated carbon

HOW TO CITE:
Jaiyeola AT, Bwapwa JK. Treatment technology for brewery wastewater in a water-scarce country: A review. S Afr J Sci. 2016;112(3/4), Art. #2015-0069, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20150069

Water is a scarce resource in many parts of the world; consequently the application of innovative strategies to treat wastewater for reuse is a priority. The brewery industry is one of the largest industrial users of water, but its effluent is characterised by high levels of organic contaminants which require remediation before reuse. Various conventional treatment methods such as anaerobic and aerobic systems, which are effective options because of their high removal efficiencies, are discussed in this study. Other methods such as membrane based technologies, carbon nanotubes, activated carbon, electrochemical methods, algal ponds and constructed wetlands are also analysed. Their efficiency as well as advantages and disadvantages are highlighted and evaluated. Combinations of various treatment processes to improve the quality of the final effluent are discussed.

Introduction

The availability of usable fresh water is a worldwide concern, but it is especially important for countries like South Africa that have both limited water resources and a steadily growing population. According to the Strategic Water Partners Network, South Africa, by the year 2030 the demand for fresh water in South Africa will exceed supply by 17% because of population growth, rapid industrialisation, mechanisation and urbanisation.[1]

For such countries, it is extremely important to develop means for reducing water consumption by industries such as the brewing industry. The production of beer on a commercial scale requires much more water than just what is contained in the beer itself if one takes into account the water used for cooling and hygienic purposes. Brewery effluent is loaded with high levels of organic matter, nutrients and solids, as shown in Table 1, which are not easy to remove using traditional methods.

An example of a brewing company that is making an effort to minimise the water requirements of the beer production process is the South African Breweries, which is now the world's second-largest beer producer after merging with Miller Brewing of the USA to become SABMiller. However, it is a very challenging task because of the high consumption of water recorded during beer production. The average usage by the SABMiller Group is reported to be around 4.6 L of water per 1 L of beer.[3] This implies that the amount of effluent discharged is greater than the amount of beer produced. SABMiller achieved absolute water reduction of 28% between 2008 and 2015 despite volume growth.[3,4]

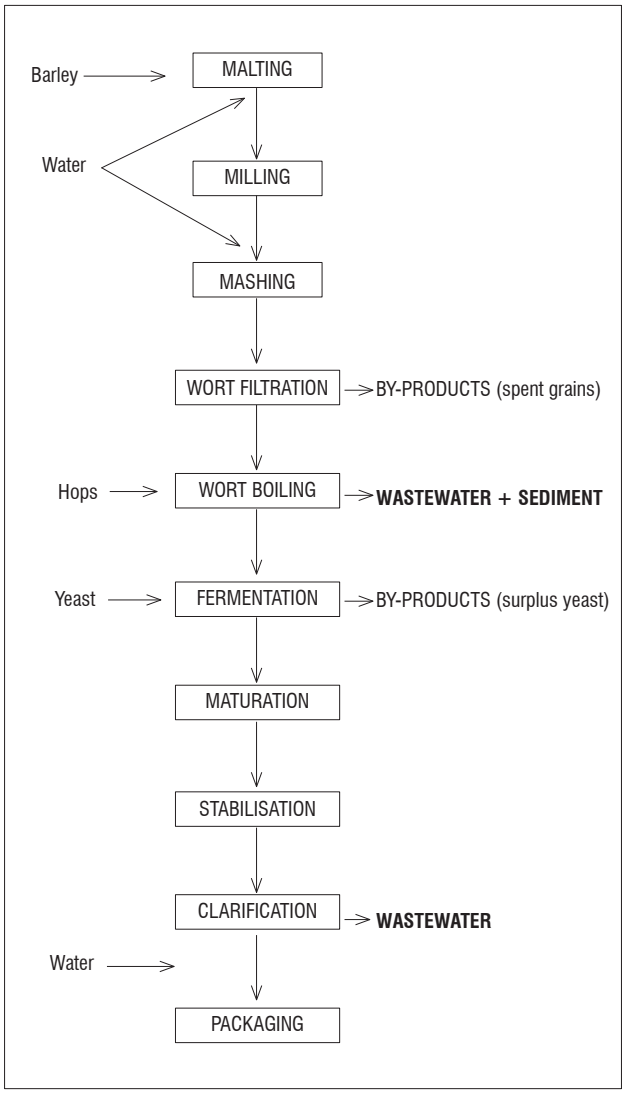
The treatment of effluent for reuse is the preferred and most widely supported methodology. Wastewater from brewery operations has a high nutrient concentration and the traditional method of disposal by delivering the water to a municipal sewage treatment plant is both wasteful and a source of concern for downstream recipients. In addition, treatment of wastewater by municipal treatment plants represents a very significant cost to brewery operators. Given the potential value of nutrient-rich brewery wastewater, and the need to conserve water usage, it therefore makes a great deal of sense to find other uses for the wastewater as well as a means of using less water overall.

Brewery operations generate large volumes of wastewater through a sequence of processes represented in Figure 1. This study analyses various conventional treatment methods for brewery wastewater and highlights the strengths and weaknesses of each method. The analysis describes the processes with technical details including some aspects related to costs, trends and achieved performance.

Table 1: Characteristics of brewery effluent

Parameter	Unit	Brewery effluent composition	Typical brewery benchmarks
Flow	Not determined	Not determined	2–8 hL effluent/hL beer
Chemical oxygen demand (COD)	mg/L	2000–6000	0.5–3 kg COD/hL beer
Biochemical oxygen demand (BOD)	mg/L	1200–3600	0.2–2 kg BOD/ hL beer
Total suspended solids (TSS)	mg/L	200–1000	0.1–0.5 kg TSS/hL beer
Temperature		18–40	
pH		4.5–12	
Nitrogen	mg/L	25–80	
Phosphorus	mg/L	10–50	
Heavy metals	mg/L	Very low	

Source: Adapted from Driessen and Vereijken[2]



Source: Adapted from Varnam and Sutherland[5]

Figure 1: Technological processes in brewery operations

Analysis of various treatment processes

Generally, the commercial beer-making process goes through the chemical and biochemical stages of mashing, boiling, fermentation and maturation, as well as separation of the wort, clarification of the wort and clarification of the rough beer product.[6] It is crucial to emphasise that these processes consume water and generate effluent streams. The solid and liquid waste fractions from these various steps, especially those containing nutrients, have been the subject of much research and have focused on either generation of revenue or at least a reduction in the cost of disposal. To separate treated water from saline water, the filtering aid known as Kieselguhr is used; however, it is entirely composed of mined diatomaceous earth, which is considered a hazardous waste. Hence, it carries significant costs associated with its use and disposal as waste. The disposal of mined diatomaceous earth is difficult because its weight increases as a result of entrained liquid. Thus, a methodology that would reduce the amount of entrained liquid and allow recovery of the Kieselguhr in a usable form would be beneficial with respect to brewery water consumption and the reduction of wastewater volume.

Another important consideration of brewery wastewater effluent is its chloride salt and ammonia content. According to a report from SABMiller regarding the performance of a pilot treatment plant initiated at Port Elizabeth in 2008, the majority of this effluent is only suitable for discharge into saline estuaries, because of the concentration of chlorides.[7] This factor inhibits the use of effluent for secondary purposes that are sensitive to salt, such as agricultural/horticultural or even reuse in the brewing industry. Historically, brewery wastewater treatment options have been limited. The classification of beer as a natural food product places strict obligations on its production processes, which includes restrictions on the input of reused materials such as wastewater. All inputs must meet the highest standards of food-grade materials, which would be complicated if the water originated from waste. Therefore, any process that promotes the reuse of brewery wastewater must generate water that meets the standards placed on fresh input water which is a costly and challenging obligation. Some of the available treatment options for brewery wastewater are described below.

Disposal, pre-treatment and treatment

To date, there have been few options available for the treatment of effluent from the brewery industry. The primary mode of treatment has been to simply dump the wastewater into the environment without any kind of treatment. However, this practice has the major drawback of environmental pollution. Pressure for beer production to satisfy increasing consumer demand generates large volumes of effluent. This has forced the beer brewing industry to implement a pre-treatment option before disposal. Generally, this pre-treatment aims to remove solids and reduce the pollutants in the water. Simple dumping has been replaced by disposal of the pre-treated effluent to municipal water resource recovery facilities, where it is added to the municipal wastewater stream for treatment before release into the environment. The mixture goes through a traditional municipal wastewater treatment process of primary, secondary and tertiary treatments, which are physico-chemical and biological treatments. The chemical oxygen demand (COD) of brewery wastewater is relatively high, generally between 2000 mg/L and 6000 mg/L, because of the presence of organic materials as by-products of the brewing process.[8] The presence of these materials represents the depletion of oxygen in the water that occurs through the oxidative chemical processes that break it down to carbon dioxide and water. The abundance of brewery wastewater in the municipal wastewater treatment systems increases the level of contaminants in the system thereby, adversely affecting the overall performance of the plant.

Anaerobic and aerobic treatment

The abundance of organic substances in brewery wastewater is the main cause of high COD which requires efficient treatment methods that can easily remove organic pollutants. Generally, biological methods are a more viable option than physico-chemical methods regarding the efficient removal of organic matter from wastewater. Therefore, treatment by passing the wastewater through an anaerobic digestion process using anaerobic bacteria and then through an aerobic process using 'activated sludge' has become the standard and most recommended process for brewery effluent. Both anaerobic and aerobic methods are widely used today, aiming to reduce the COD of brewery effluent before it is transported to municipal water resource recovery facilities.

Haydon[9] indicated that anaerobic digestion of brewery effluent is increasingly being utilised to generate an energy source such as biogas. This requires the hydrogen sulfide content of the biogas to be scrubbed out. The generated biogas can be used to maintain the operating temperature of the anaerobic digestion system or to generate revenue. As shown in Table 2, aerobic systems are more efficient in terms of removal of organic pollutants but require the use of oxygen which increases operating and capital costs. Both anaerobic and aerobic processes require high capital costs, however, they differ on operating costs which are lower for anaerobic compared to aerobic processes. The start-up cost of anaerobic digestion is high because of the laborious seeding process required for an appropriate culture of microorganisms and anaerobic processes are unable to efficiently reduce the level of some nutrients such as nitrates.

In the search for efficient brewery effluent treatment methods, the treatment of brewery effluents has been able to draw from the experience of other applications and industries, particularly the treatment of acid rock drainage through the use of aerobic and anaerobic digestion structures.

Table 2: Comparison between anaerobic and aerobic systems

	Anaerobic	Aerobic
C Chemical oxygen demand removal	65–90%	90–98%
Energy production	High: CH_4 is produced as biogas	Low: only CO_2 released
Energy consumption	Low	High
Sludge production	Low: high solid retention	High
Nutrients (N/P) removal	Low	High
Space requirement	Low	High
Discontinuous operation	Easy	Challenging

Source: Adapted from Driessen and Vereijken[2]

Aerobic digestion structures are typically wetlands and shallow settling ponds that provide an oxidative environment for entrained contaminants, while anaerobic digestion structures use anaerobic bacteria to digest many of the nutrients that are found in brewery wastewater. These methods are separate from the anaerobic digestion and activated sludge processes used to treat brewery wastewater.

Cowan and Rende,[10] discusses the generation and use of biogas as a self-supporting aspect of the Integrated Algae Ponding System. This system uses both anaerobic and aerobic processes in a deep fermentation pit. In the lower section of the pit, anaerobic digestion consumes organic matter in wastewater. The odoriferous materials are oxidised through consumption by aerobic algae, driven by solar input, in the upper section of the pit. Similar to other systems, it is common practice in this system to include a fish farming operation at the last stage of the process before release into the environment. The fish farm is added as a means of 'polishing' the effluent water, as the fish consume any remaining algae. This aspect of the process is problematic for the South African brewing industry, as suitable indigenous fish species have not been identified. Also, the number of the fish farms that have been set up reduced quickly as people took advantage of the novel source of food. Consequently, the quality of effluent discharged into the environment is not as 'polished' as might be expected.

High rate algal ponds and constructed wetlands treatment

Jones et al.[11] completed a study on SABMiller's Eden Project carried out at its Ibhayi brewery location. This was an experimental facility aiming at testing and adapting the remediation of brewery wastewater using high rate algal ponds and constructed wetlands (HIRAP/CW) technology. The experimental station used a specially designed greenhouse to test the use of the treated effluent for its suitability in secondary applications. These experiments sought to reduce or eliminate the chloride content of the wastewater. The treated effluent proved to be suitable for the growth of vegetables hydroponically, as well as for raising fish. Additionally, sodium tolerant crops such as bananas were grown successfully with the treated effluent. The SAB Miller project needs to be expanded in such a way that all effluent from the Ibhayi brewery and SAB Miller's other locations can be treated in a similar manner.

The SAB Miller experiment is unique and represents the first case study of a commercial brewery treating brewery effluent in South Africa. The method proved to be effective in reducing both ammonia and phosphate levels in the effluent water to regulated standards. However, the method is unable to remove chlorides and sodium salts from the effluent. Hence, the effluent remains only suitable for discharge into a saline estuary. The advantage of this technology is its extremely low cost and reliability. The technology is adapted from the remediation of acid mine drainage and uses both a bacterial phase to reduce organic components in the wastewater and an aerobic algal pond phase to consume mineral components such as phosphates. Kivaisi[12] and Shrestha et al.[13] examined the use of constructed wetlands (CW) technology for the reduction of COD in the treatment of municipal wastewater. They reported that it could be reduced by 90–94% by a 'mature' wetland. Similarly, Shepherd et al.[14] examined the use of CW technology in the treatment of winery wastewater, which is similar to brewery wastewater, and reported a COD removal efficiency of up to 98%. All this evidence showed that the application of HIRAP/CW technology for the treatment of brewery effluent could be eminently successful for reducing COD in effluent before passing it to further treatment or to the environment. However, as mentioned earlier, the saline content of the effluent was unaffected by HIRAP/CW treatment.

Carbon nanotubes based treatment

Simate[15] used carbon nanotubes to treat effluent from the brewing industry. The carbon nanotubes were prepared from carbon dioxide (CO_2), which is another important by-product of the beer brewing industry. The nanotubes were found to be very effective as a flocculating agent and as a granular filtration medium during treatment. It was reported that treatment using carbon nanotubes reduced the turbidity of the effluent to less than 5 NTU and removed 96.% of the effluent COD. Given the current attention to carbon nanotubes/graphene technology, this is a promising alternative for the treatment of brewery effluents in the near future. The strength of this technology is the fact that carbon nanotubes can be produced from purified CO_2 as a carbon source. Furthermore, it is a possible means of using up large amounts of the CO_2 produced by the brewing industry from the fermentation process. Therefore, this could contribute to the attenuation of global warming effects from the brewing industry. However, carbon nanotube treatment for effluent from the brewing industry is not yet a viable option. The technology requires large amounts of energy for CO_2 purification and carbon nanotube synthesis. Therefore, the costs of scaling up to a commercial treatment operation would be exorbitant. Also, the unavoidable release of carbon nanotubes into the environment might result in massive environmental and health effects with unexpected consequences. Carbon nanotubes and other fullerene-type compounds are naturally occurring products in soot from carbon combustion, while graphene is the basic molecular building block of graphite and coal. The challenge with this carbon nanotubes technology is the fact that the environment has never been exposed to it on a large scale. For this reason, more investigations are required on capital and operating costs as well as environmental impact and process safety of carbon nanotubes.

Advanced oxidation treatment process

Advanced oxidation treatment processes (AOP) are widely used in wastewater treatment, especially in alcohol distilleries, which generate almost the same type of effluent as the brewing industry, with high levels of organic compounds. The production process in distilleries is almost same as breweries because they both involve fermentation.[16] Ozone, hydrogen peroxide and ultraviolet irradiation are used to produce hydroxyl radicals (•OH) during the first stage of the oxidation. In the second stage, hydroxyl radicals (•OH) react with organic contaminants to produce precipitates. Established AOP technologies can be made possible with the help of the following combinations: ozone/hydrogen peroxide (O_3/H_2O_2), ozone/ultraviolet irradiation (O_3/UV) and hydrogen peroxide/ultraviolet irradiation (H_2O_2/UV).[17] Bes-Piá[17] revealed that both

ozone and hydroxyl radicals (•OH) are strong oxidants and are capable of oxidising a number of organic compounds. Ozone is a powerful oxidant that reacts with a great number of organic compounds and facilitates the removal of organic pollutants from wastewater, once dissolved in water. It acts in two different ways, (1) by direct oxidation as molecular ozone and (2) by indirect reaction through formation of secondary oxidants like free radical species, particularly the hydroxyl radicals (•OH).

Pala and Erden[18] reported another AOP process known as Fenton's oxidation using Fenton's reagent, which is a mixture of hydrogen peroxide and iron salts (Fe^{2+} or Fe^{3+}). This technology is based on the production of hydroxyl radicals (•OH) that ultimately leads to precipitation or decolourisation of the effluent.[18] Furthermore, Fenton produces homogeneous reaction and is environmentally friendly.[18] There are other AOP processes such as catalytic oxidation using the combination of TiO_2/UV, catalytic ozonation and boron doped diamond electrodes; however, they are still only on a laboratory scale. Oxidation processes show a promising future for their application in many wastewater treatment projects. They are emerging technologies with great potential to remove pollutants from many types of wastewater including brewery effluent. These processes utilise the strong oxidising power of hydroxyl radicals to oxidise organic compounds to the preferred end products of carbon dioxide and water. However, sometimes they can be costly because supplementary treatment may be necessary to remove ozone. In addition, there is a problem of turbidity/NO_3 interference to resolve.

Membrane filtration treatment

Membrane filtration can be an effective treatment depending on the type of effluent. This method can achieve up to 99% COD, Biochemical oxygen demand (BOD), and total suspended solids (TSS) removal. If the required final product is potable water, reverse osmosis can be added to the process. Microfiltration (MF), Ultrafiltration (UF) and Nanofiltration (NF) membranes have been used successfully to remediate brewery effluent.[4,19,20] The challenge with membrane technologies is fouling and energy consumption, but new types of membranes with anti-fouling properties make the membrane process a viable treatment option. More studies are required in this area to produce membranes that can work with minimum fouling and efficient use of energy. Conversely, membrane technologies should only be used as a polishing step after a pre-treatment option involving anaerobic or aerobic processes or a combined anaerobic/aerobic process.

Daufin et al.[21] reported that in the brewery industry, cross-flow or dynamic filtration can play a significant role that can be a technological alternative to the conventional solid and liquid separations. The efficiency of this method depends on two factors (1) the recovery of extract during the wort clarification, and (2) beer recovery from tank bottoms, more especially fermentation and maturation vessels.[21] Currently, tank bottom recovery is reported to be the principal membrane application in brewing.[21] MF can be utilised for three purposes: mash separation, clarification of rough beer and cold sterilisation of clarified beer before conditioning. UF and NF are suggested for effluent treatment; however, this does not exclude pre-treatment processes. It is important to note that industrial applications are more focused on the clarification of rough beer and sterile filtration of clarified beer.

Membrane bioreactor treatment

Membrane bioreactor (MBR) treatment is a successful technology for wastewater treatment and during the last decade has also produced successful results for drinking water.[22,23] It is a combination of two proven technologies: enhanced biological treatment using activated sludge or an anaerobic unit, and membrane filtration. The development of various combinations of membranes with other conventional treatment components is justified by the increasing water price and continuous depletion of water resources. Therefore, MBR is seen as an economical and technically viable option for wastewater treatment.[24] An MBR system is constructed in such a way that a membrane is integrated with a bioreactor. Two MBR process configurations can be identified: side-stream and submerged. In a side-stream process, the membrane module is placed outside the reactor and the reactor mixed liquor flows over a recirculation loop containing the membrane. In a submerged configuration, the membrane is placed inside the reactor and submerged into the mixed liquor. Side-stream MBRs are more energy intensive than submerged MBRs because of high operational transmembrane pressures and the significant volumetric flow required to achieve the desired cross flow velocity.[25,26] Submerged MBRs use more membrane area and operate at lower flux levels.[25]

Dai et al.[27] reported that MBR technology can be applied successfully to brewery wastewater treatment and showed COD removal of up to 96% using an upflow anaerobic sludge blanket (UASB) reactor with an integrated membrane. Brewery effluent treatment using MBR was also reported in other studies.[28,29,30] In most of these studies, significant amounts of COD removal of up to 90% were recorded which indicates that the MBR process can be a successful option for the treatment and reuse of brewery effluent. However, as with any membrane process, fouling is the greatest challenge and needs to be managed with routine cleaning and maintenance. Energy consumption can also be a weakness especially when it comes to side stream membranes. Capital costs can also be higher because of the combination of two units: membrane, and anaerobic or aerobic reactor.

Microalgae based treatment

Mata et al.[31] analysed the potential of using a microalgae strain known as *Scenedesmus obliquus* for brewery effluent treatment. Usually, microalgae can grow in the brewery effluent using contaminants as nutrients. According to Mata et al.[31], the best operating conditions are aerated cultures exposed to high intensity light for 12 h daily. It was reported that maximum biomass growth was achieved after 9 days with an output of 0.9 g of dry biomass per litre of culture. Removal of contaminants with 57.5% COD and 20.8% total N_2 was recorded after 14 days. The final values of COD and total nitrogen were found to be 1692 mg O_2L and 47 mg N_2/L respectively. Compared to discharge standards, which are 150 mg O_2/L and 15 mg N_2/IL, respectively, final values of COD and total nitrogen were higher. As a result, wastewater treatment by microalgae can be achieved in either the secondary or a tertiary phase of the treatment, combined with other treatments. Membrane technologies can easily be added, in this case, for the polishing step. After utilisation, the resulting algal biomass can be collected and used for biofuel production. Algae treatment for wastewater is an emerging method that can also help to solve the challenge of brewery effluent treatment at lower costs. More investigations are required to find suitable strains that are able to efficiently remove large amounts of organic contaminants from brewery effluents, but there is a potential and promising future regarding this type of treatment.

Treatment using air cathode microbial fuel cells

Feng et al.[32] examined the efficiency of a microbial fuel cell (MFC) method to determine its suitability for brewery effluent treatment. This method is still new and has drawn worldwide interest because it can generate electricity from organic matter present in wastewater. MFCs are devices that use microorganisms to convert chemical energy into electricity from organic matter. MFC is a combined system with anaerobic and aerobic characteristics. Anaerobic digestion by microorganisms takes place in the solution close to the anode with the cathode exposed to the oxygen. Electrons released by bacterial oxidation of organic compounds are transferred to the cathode where they react with oxygen from water. Furthermore, Feng et al.[32] reported that the effectiveness of MFC treatment would require a good understanding of how solution chemistry and operational parameters affect the efficiency of the treatment. Feng et al.[32] used parameters such as maximum power densities and COD removal as functions of temperature, effluent strength, and coulombic efficiencies (CE) to test the efficiency of MFC. It was found that when temperature was decreased from 30 °C to 20 °C, the maximum power density was reduced from 205 mW/m^2 to 170 mW/m^2. However, COD removals and CE decreased slightly with temperature decrease. Also, the performance of the reactor was strongly affected by the buffering capacity. Power density was significantly increased by 136% with the addition of 50 mM phosphate. COD removal efficiency was 85% and 87% at 20 °C and 30 °C, respectively. Performance of sequential anode-cathode MFC achieved

COD removal efficiency of more than 90%. In another study, Mathuriya et al.[33] achieved a COD removal of up to 94% using the MFC method for brewery wastewater. In conclusion, MFC more specially the sequential anode-cathode, can provide a new approach for brewery wastewater treatment while offering an alternative method of generating energy.

Activated carbon based treatment

Activated carbon is one of the most powerful adsorbents for removing a wide range of contaminants from industrial and municipal wastewater, landfill leachate and contaminated groundwater. This powerful adsorbent can cope with a wide range of contaminants including organic contaminants present in brewery effluent. Different contaminants may be present in the same discharge and carbon may be used to treat the total flow, or it may be utilised to remove specific contaminants as part of a multistage approach.[8] Brewery effluent is characterised by its distinctive odour resulting from fermentation and other beer-making processes and it may also contain carbon-sulfur bonds. Molecules with carbon-sulfur bonds and aromatic rings often smell and produce a bad taste, but these are molecules preferentially adsorbed on carbon. Carbon's de-chlorinating capability results from its ability to act as a reducing agent that reacts with strong oxidising agents such as chlorine dioxide and hypochlorous acid. Carbon adsorption is used in the brewing process to treat tannic acid for odour removal.[8] Carbon is also used to remove colour from malts for use in clear beers and other flavoured malt beverages. Activated carbons are an effective treatment option to assure that treated water is taste and smell free. It is the least expensive process and does not require electricity or high water pressure. It reduces a wide variety of organic contaminants and can be designed to reduce levels of some inorganic chemicals like lead and arsenic. However, activated carbon is ineffective against many inorganic contaminants such salts, iron, fluoride, aluminium and calcium.[8]

Treatment of brewery effluent using electrochemical methods

The electrochemical method of wastewater treatment was first used to treat sewage generated on board ships.[34] Later, the application of electrochemical treatment was widely used in industrial wastewater treatment that is rich in refractory organics and chloride content.[35] Vijayaraghavan et al.[36] developed a novel brewery wastewater treatment method based on in situ hypochlorous acid generation. The generated hypochlorous acid served as an oxidising agent that destroyed organic compounds present in the brewery wastewater. COD removal of up to 97% was achieved in this study. The hypochlorous acid was generated using a graphite anode and stainless steel sheet as a cathode in an undivided electrolytic reactor. Initially, during electrolysis, chlorine was produced at the anode and hydrogen gas at the cathode. This method is appropriate for the degradation of biorefractory organic contaminants because complete or partial decomposition of organic substances is achievable. The advantage of electrochemical methods is that they require less hydraulic retention time and are not subject to failure resulting from variation in wastewater strength or the presence of toxic substances.

Non-thermal quenched plasma technology treatment

Plasma usually results from the increased energy of a gas provided by various sources, such as electric, magnetic, shock waves, ultrasound, thermal and optical (laser) sources.[37] Plasma is a highly ionised gas that occurs at high temperatures and is considered as a fourth state of matter because intermolecular forces created by ionic attractions and repulsions give these compositions distinct properties.[38] Plasma is similar to gas in terms of physical properties as it does not have fixed shape or volume. However, it differs from gas by being able to form structures such as filaments, beams and double layers under the influence of a magnetic field. Doubla et al.[39] demonstrated the use of this technology for the treatment of brewery effluent.[39] Humid air plasma created by an electric gliding arc discharged in humid air was used during the experiment and contributed to the removal of organic pollutants from brewery effluent. The gliding arc discharge in humid air was able to produce •NO and •OH radicals with strong oxidising characteristics. The •OH radical is a very powerful oxidising agent that is responsible for oxidation reactions with organic contaminants considered as targets during treatment. The study recorded BOD removal efficiencies of 74% and 98%. The technology is also effective in the neutralisation process of alkaline effluents because of the pH lowering effect of the plasma treatment originating from the production of nitrate ions. This method can be combined with biological treatments to further remove organic pollutants easily and quickly to an acceptable level for effluent reuse.[39] Although high removal efficiencies can be achieved, the method is expensive because of the high energy requirements of the gas and energy sources such as laser. If combined with biological treatment as mentioned earlier, capital and operating costs will increase.

Discussion

Brewery effluent still needs efficient solutions for remediation that are low cost. Table 3 shows the performance of the various methods discussed in this study in terms of COD removal. These methods show promise but each have weaknesses that need to be addressed for a better outcome. As the demand for beer and other brewery products increases, so does the amount of generated effluent. The brewing industry is focused on producing more beer but most brewers are ignoring the optimisation of water used in the process and the development of an efficient treatment for the generated effluent. Generally, brewery effluent is simply dumped into the nearest municipal water treatment plant for processing along with sewage and other wastes.

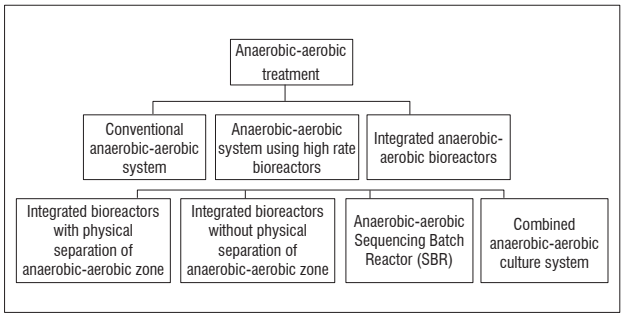
Table 3: Methods used for the treatment of brewery effluent and the efficient removal of chemical oxygen demands

Process	COD removal %
Quenched plasma	74–98[39]
Upflow anaerobic sludge blanket reactor	73–91[40]
Aerobic reactor	90–98[2]
Combined bioreactor	98
Membrane bioreactor	96[27]
Electrochemical method	97[34]
Microbial fuel cells	94[41]
Nanofiltration	96[20]
Reverse osmosis	100[42]

Source: Adapted from Simate et al.[8]

Some efforts have been made in the recovery of economic value from the solid organic residues of the brewing industry, such as the sale of malting and yeast residues for incorporation into animal feedstock. However, these efforts do not have any impact on water consumption and wastewater remediation in the brewery environment. According to Haydon[9], the brewing industry has long been aware of possible uses for brewery effluents, such as the generation of biogas, but has not pursued this option to date because of their lack of economic viability. However, this appears to be changing as the rising costs of input resources make utilisation of by-products economically more feasible. The biggest challenge for the brewing industry remains the volume of wastewater that it produces. Shepherd[14] reported that the consumption of water by brewing industry is almost 10 L per 1 L of beer, which is double the figure reported in the study by Fillaudeau et al.[6] An efficient brewery should use between 4 L and 7 L of water to produce 1 L of beer.[43] Kanagachandran[44] estimated that about 3–10 L of effluent is produced per 1 L of beer produced in breweries. These figures stress the importance of reducing effluent volumes in the industry and also imply that the quantity of generated brewery effluent will depend on

the management of the production process. More studies need to be undertaken to optimise the use of water in the brewing industry to reduce the amount of effluent generated during production. The brewery effluent is rich in nutrients, which represent a resource that could be developed into a valuable commodity. For example, The Eden Project undertaken by SABMiller pointed the way to remediation of brewery effluent that could lead to its complete reuse within the brewing industry.[45] Complete reuse would cut the water requirements of the brewing industry by a huge amount, given that 70% of water used by the industry is lost as waste. In one of SABMiller's breweries, this would amount to the recovery of 7 million m^3 (7 billion litres) from one single brewery plant per year. The obstacle that needs to be overcome in these efforts is the generation of an effluent that is eventually close in quality to the water used as an input stream into the brewing process. To reach that stage, the nutrient and the mineral content must be reduced. Particularly salts, ammonia, phosphate and other inorganic materials must be reduced to below those of the input water. At the present state, the effluent contains high levels of salts, which renders it not viable for use in secondary purposes. Despite the fact that algal pond treatment of brewery effluent is an acceptable start, there is still room for new methods. Further improvements are required to implement treatment methods other than anaerobic digestion and algal ponds, possibly a combined aerobic and anaerobic system could be a suitable option as illustrated in Figure 2.[46]



Source: Driessen et al.[46]

Figure 2: Combined aerobic and anaerobic systems used in the treatment of wastewater.

Figure 2 shows that there are basically four types of integrated aerobic-anaerobic bioreactors and the properties of these integrated bioreactors as outlined by Driessen et al.[46] are: (1) most of the COD (70%–85%) from the anaerobic reactor is converted into biogas, (2) almost 98% of the COD and nutrients are removed in the anoxic/aerobic post-treatment stage. The advantages of these combined treatment technologies over aerobic conventional treatment processes is that there is a reduction in the production of sludge (bio), less space is required and a positive energy balance is maintained.[2] Recently, the combination of tall and slender internal circulation anaerobic reactors with airlift reactors (aerobic reactors) has resulted in the design of well compacted effluent treatment plants that meet the required surface water quality.[2] With this method it is possible to successfully treat brewery effluent, as illustrated in Table 3, which shows that recorded COD removals are between 73% and 100%. These values are very high and there is no questioning their effectiveness when looking at the methodological approach and equipment used to achieve such results. Some methods are suitable for pre-treatment and others can be used as a polishing step. This supports the option of combining processes to achieve the high quality required for rendering effluent suitable for reuse in the brewing process. However, in these studies, the economic aspects are not always meticulously analysed, which makes it difficult to confirm with certainty if a method or a combination of methods would be economically viable for treating brewery effluent. There is an absence of adequate economic analysis on operating and capital costs to help make an effective choice for treatment, especially when dealing with large volumes of effluent.

Conclusions and recommendations

Researchers must continue to discover effective ways to reduce the amount of brewery effluent that is produced annually. Improvements must be found to recover brewery effluent as much as possible so that it can be reused. Water scarce countries like South Africa, and its brewing industry, face a crisis of clean water availability. While new technologies such as the application of carbon nanotubes appear to have the potential to achieve highly desirable results in laboratory tests, the technology is not even close to being ready for application in the brewery industry and would be enormously expensive on a large scale. In addition, there are the unidentified health risks of carbon nanotubes presence in the environment. Similarly, the 'low technology' approach for remediation is not capable of achieving the desired results. Biological methods such as aerobic and anaerobic treatments are widely used because of their capacity to remove organic contaminants, achieving high COD removal, but are characterised by high capital or operating costs. These methods are used as pre-treatment options and further treatments may be required.

This review has shown some promising results from MBR, quenched plasma, MFC and the electrochemical method. There is a great potential in these methods; however, the technology costs resulting from energy consumption and maintenance may be inhibitory. The Algal ponds and constructed wetland method is reliable and has the advantage of lower costs. This method shows effectiveness in the removal of ammonia and phosphates but it is unable to remove chlorides and sodium salts. Membrane filtration is being used for industrial wastewater treatment including brewery effluent. Membrane technology has found application in drinking water and wastewater reuse and has undergone rapid development and improvement in quality and costs over the last decade, and could be used as a polishing step after pre-treatment options. Activated carbon based treatment methods can be an appropriate treatment option for the brewery industry. It is a cheap option that could easily allow the removal of organic contaminants from brewery effluents. However, it may be challenged by health and environmental concerns related to the use of carbon or coal for the treatment of large amounts of effluent.

The integration of technologies or processes can also be an option to improve the quality of the final product. This option needs more investigation, especially regarding operating and capital costs including energy consumption, maintenance, process efficiency, water consumption and optimisation. It is incumbent upon the brewery industry to invest in developing alternatives for treatment of brewery effluent, looking at the effluent as a commodity resource rather than as waste. Breweries around the world would do well to form a research consortium to address the problem of brewery wastewater, its remediation and reuse. While the largest breweries are located in the wealthiest economic regions of the world, they do not face the eventual water shortages of less water-affluent countries like South Africa. The growth of breweries in developing countries demands that water conservation and remediation methods in the brewery industry be developed sooner rather than later. It is further recommended that a thorough or a more detailed cost analysis of the various treatment processes in this review be carried out to help determine the most cost-effective way to treat wastewater from breweries.

Authors' Contributions

The authors worked together from topic development, collection of information, discussion, critical analyses, writing of the manuscript up to the final corrections. The article is a result of synergy of both authors.

References

1. Department of Water Affairs. Strategic Water Partners Network, South Africa: Closing the water gap by 2030. Johannesburg: NEPD Business Foundation

2. Driessen W, Vereijken T. Recent developments in biological treatment of brewery effluent. The Institute and Guild of Brewing Convention; 2003 March 2–7; Livingstone, Zambia. The Netherlands: Paques Water Systems; 2003. p. 10.

3. SAB Miller Group. SAB Miller position paper – water: Make more beer but use less water [paper on the Internet]. c2009 [cited 2016 Feb 06]. Available from: http://www.sabmiller.com/docs/default-source/sustainability-documents/position-paper-water.pdf?sfvrsn=2

4. SAB Miller PLC. Annual Report 2015 [report on the Internet]. c2015 [cited 2016 Feb 06]. Available from: http://www.sabmiller.com/docs/default-source/investor-documents/financing-documents/annual-report-2015.pdf?sfvrsn=2

5. Varnam AH, Sutherland JP. Alcoholic beverages: Vol 1 Beer. In beverages – technology, chemistry and microbiology, Vol 2. Food products series. London: Chapman & Hall; 1994. p. 296–261.

6. Fillaudeau L, Blanpain-Avet P, Daufin G. Water, wastewater and waste management in brewing industries. J Clean Prod. 2006;14:463–471. http://dx.doi.org/10.1016/j.jclepro.2005.01.002

7. Engineering News. Wastewater treatment pilot making green brewing strides at PE in 2008. Engineering News. 2011 April 15. Available from: http://www.engineeringnews.co.za/article/environmentally-friendly-plant-treats-brewery-effluent-2011-04-15

8. Simate GS, Cluett J, Iyuke SE, Musapatika ET, Ndlovu S, Walubita LF. The treatment of brewery wastewater for reuse: State of the art. Desalination. 2011;273(2–3):235–247. http://dx.doi.org/10.1016/j.desal.2011.02.035

9. Haydon P. The Boom in Biogas. Brewers' Guardian. 2011 Sept/Oct;37–40. Available from: http://www.brewersguardian.com/features/biogas-production-waste-water-brewing.html

10. Cowan AK, Rende DS. Integrated algae ponding system: Technical description. Grahamstown: Institute for Environmental Biotechnology, Rhodes University; 2012. Available from: https://www.ru.ac.za/media/rhodesuniversity/content/ebru/documents/IAPS%20Technical%20Description.pdf.pdf

11. Jones CLW, Britz P, Davies MTT, Scheepers R, Cilliers A, Crous L, et al. 2011. The wealth in brewery effluent - Water and nutrient recovery using alternative technologies. Fifteenth International Water Technology Conference, IWTC-15; 2011 May 28–30; Alexandria, Egypt.

12. Kivaisi AK. The potential for constructed wetlands for wastewater treatment and reuse in developing countries: A review. Ecol Eng. 2001;16:545–560. http://dx.doi.org/10.1016/S0925-8574(00)00113-0

13. Shrestha RR, Haberl R, Laber J, Manadhar R, Mader J. Application of constructed wetlands for wastewater in Nepal. Water Sci Technol. 2001;44:381–386.

14. Shepherd HL, Grismer ME, Tchobanoglous GG. Treatment of high strength winery wastewater using a subsurface-flow constructed wetland. Water Environ Res. 2001;73:394–403. http://dx.doi.org/10.2175/106143001X139434

15. Simate G. The treatment of brewery wastewater using carbon nanotubes synthesized from carbon dioxide carbon source [PhD Thesis]. Johannesburg: University of the Witwatersrand; 2012. http://dx.doi.org/10.1016/j.watres.2011.12.023

16. Mohana S, Acharya BK, Madamwar D. Distillery spent wash: Treatment technologies and potential applications. J Hazard Mater. 2009;163(1):12–25. http://dx.doi.org/10.1016/j.jhazmat.2008.06.079

17. Bes-Piá A, Mendoza-Roca JA, Roig-Alcover L, Iborra-Clar A, Iborra-Cla MI, Alcaina-Miranda MI. Comparison between nanofiltration and ozonation of biologically treated textile wastewater for its reuse in the industry. Desalination. 2003;157:81–86. http://dx.doi.org/10.1016/S0011-9164(03)00386-2

18. Pala A, Erden G. Decolorization of a baker's yeast industry effluent by Fenton's oxidation. J Hazard Mater. 2005;127:141–148. http://dx.doi.org/10.1016/j.jhazmat.2005.06.033

19. Ince BK, Ince O, Sallis PJ, Anderson GK. Inert COD production in a membrane anaerobic reactor treating brewery wastewater. Water Res. 2000;34:3943–3948. http://dx.doi.org/10.1016/S0043-1354(00)00170-6

20. Braeken L, Van Der Bruggen B, Vandecasteele C. Regeneration of brewery waste water using nanofiltration. Water Res. 2004;38:3075–3082. http://dx.doi.org/10.1016/j.watres.2004.03.028

21. Daufin G, Escudier J-P, Carrère H, Bérot S, Fillaudeau L, Decloux M.. Recent and emerging applications of membrane processes in the food and dairy industry. Food Bioprod Process. 2001;79:89–102. http://dx.doi.org/10.1205/096030801750286131

22. Li X, Chu, HP. Membrane bioreactor for the drinking water treatment of polluted surface water supplies. Water Res. 2003;37:4781–4791. http://dx.doi.org/10.1016/S0043-1354(03)00424-X

23. Fan F, Zhou H. Interrelated effects of aeration and mixed liquor fractions on membrane fouling for submerged membrane bioreactor processes in wastewater treatment. Environ Sci Technol. 2007;41:2523–2528. http://dx.doi.org/10.1021/es062035q

24. Visvanathan C, Pokhrel D. Role of membrane bioreactors in environmental engineering applications. In: Roussos S, Soccol CR, Pandey A, Augur, C, editors. New horizons in biotechnology. Dordrecht: Kluwer Academic Publishers; 2003. http://dx.doi.org/10.1007/978-94-017-0203-4_21

25. Jeison D. Anaerobic membrane bioreactors for wastewater treatment: Feasibility and potential applications [PhD Thesis]. Wageningen: Wageningen University; 2007.

26. Seneviratne MA. Practical approach to water conservation for commercial and industrial facilities. City East: Queensland Water Commission; 2006. http://dx.doi.org/10.1016/B978-185617489-3.50011-4

27. Dai H, Yang X, Dong T, Ke Y, Wang T.. Engineering application of MBR process to the treatment of beer brewing wastewater. Mod Appl Sci. 2010;4(9):103–109. http://dx.doi.org/10.5539/mas.v4n9p103

28. Kimura S. Japan's aqua renaissance '90 project. Water Sci Technol. 1991;23(7–9):1573–1582.

29. Nagano A, Arikawa E, Kobayashi H. The treatment of liquor wastewater containing high-strength suspended solids by membrane bioreactor system. Water Sci Technol. 1992;26(3–4):887–895.

30. Fakhru'l-Razi A. Ultrafiltration membrane separation for anaerobic wastewater treatment. Water Sci Technol. 1994;30(12):321–327.

31. Mata TM, Melo AC, Simões M, Caetano NS. Parametric study of a brewery effluent treatment by microalgae *Scenedesmus obliquus*. Bioresource Technol. 2012;107:151–158. http://dx.doi.org/10.1016/j.biortech.2011.12.109

32. Feng Y, Wang X, Logan BE, Lee H. Brewery wastewater treatment using air-cathode microbial fuel cells. Appl Microbiol Biot. 2008;78:873–880. http://dx.doi.org/10.1007/s00253-008-1360-2

33. Mathuriya AS, Sharma VN. Treatment of brewery wastewater and production of electricity through microbial fuel cell technology. Int J Biotech Bioch. 2010;6(1):71–80.

34. Bockris JOH. Environmental Chemistry. New York: Plenum Press; 1977. http://dx.doi.org/10.1007/978-1-4615-6921-3

35. Barrera-Díaz C, Linares-Hernández I, Roa-Morales G, Bilyeu B, Balderas-Hernández P. Removal of biorefractory compounds in industrial wastewater by chemical and electrochemical pretreatments. Ind Eng Chem Res. 2009;48:1253–1258. http://dx.doi.org/10.1021/ie800560n

36. Vijayaraghavan K, Ahmad D, Lesa R. Electrolytic treatment of beer brewery wastewater. Ind Eng Chem Res. 2006;45:6854–6859. http://dx.doi.org/10.1021/ie0604371

37. Benstaali B, Moussa D, Addou A, Brisset JL. Plasma treatment of aqueous solutes: Some chemical properties of a gliding arc in humid air. Eur Phys J-Appl Phys. 1998;4:171–179. http://dx.doi.org/10.1051/epjap:1998258

38. Sutton AP. Electronic structure of materials. Oxford: Clarendon Press; 1993.

39. Doubla A, Laminsi S, Nzali S, Njoyim E, Kamsu-Kom J, Brisset JL. Organic pollutants abatement and biodecontamination of brewery effluents by a non-thermal quenched plasma at atmospheric pressure. Chemosphere. 2007;69:332–337. http://dx.doi.org/10.1016/j.chemosphere.2007.04.007

40. Cronin C, Lo KV. Anaerobic treatment of brewery wastewater using UASB reactors seeded with activated sludge. Bioresource Technol. 1998;64:33–38 http://dx.doi.org/10.1016/S0960-8524(97)00154-5

41. Wang X, Feng YJ, Lee H. Electricity production from beer brewery wastewater using single chamber microbial fuel cell. Water Sci Technol. 2008;57:1117–1121. http://dx.doi.org/10.2166/wst.2008.064

42. Madaeni SS, Mansourpanah Y. Screening membranes for COD removal from dilute wastewater. Desalination. 2006;197:23–32. http://dx.doi.org/10.1016/j.desal.2006.01.015

43. Environment Canada (EC). Technical pollution prevention guide for brewery and wine operations in the Lower Fraser Basin. ECDOE, FRAP 97-20. Vancouver: EC. 1997; p. 97–20.

44. Kanagachandran K, Jayaratne R. Utilization potential of brewery waste water sludge as an organic fertilizer. J I Brewing. 2006;112(2):92–96. http://dx.doi.org/10.1002/j.2050-0416.2006.tb00236.x

45. Fakoya MB, Van der Poll HM. Integrating ERP and MFCA systems for improved waste-reduction in a brewery in South Africa. J Clean Prod. 2013;40:136–140. http://dx.doi.org/10.1016/j.jclepro.2012.09.013

46. Driessen W, Yspeert P, Yspeert Y, Vereijken T. Compact combined anaerobic and aerobic process for the treatment of industrial effluent. Environmental Forum, Columbia–Canada: Solutions to Environmental Problems in Latin America; 2000 May 24–26; Cartegena de Indias, Columbia. The Netherlands: Paques Water Systems.

Use and usefulness of measures of marine endemicity in South Africa

AUTHORS:
Charles L. Griffiths[1]
Tamara B. Robinson[2]

AFFILIATIONS:
[1]Department of Biological Sciences and Marine Research Institute, University of Cape Town, Cape Town, South Africa
[2]Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Charles Griffiths

EMAIL:
Charles.Griffiths@uct.ac.za

POSTAL ADDRESS:
Department of Biological Sciences and Marine Research Institute, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa

DATES:
Received: 07 July 2015
Revised: 02 Oct. 2015
Accepted: 08 Oct. 2015

KEYWORDS:
Endemic species; range restriction; biodiversity; marine conservation; taxonomy

HOW TO CITE:
Griffiths CL, Robinson TB. Use and usefulness of measures of marine endemicity in South Africa. S Afr J Sci. 2016;112(3/4), Art. #2015-0249, 7 pages. http://dx.doi.org/10.17159/sajs.2016/20150249

Numerous authors have cited numbers, or proportions, of endemic species within South(ern) African marine taxa, but comparisons between these statistics are confounded by differing definitions of regional boundaries and differences among data sets analysed. These have resulted in considerable variations in published endemicity data, even within the same taxonomic group. We tabulated and compared key endemicity statistics for regional marine taxa and explained biases in the data sets. The most comprehensive data sets available give overall marine endemicity within the national boundaries of South Africa as 28–33%, but estimates within individual taxa making up these totals vary enormously, from 0% (Aves, Mammalia) to over 90% (Polyplacophora). We also examined published data documenting localised endemicity patterns around the coastline. These consistently show the highest numbers of endemics occurring along the South Coast. There are logical biogeographical reasons to expect this trend, but endemicity rates are also inherently biased by distance from defined political boundaries and by differing sampling effort locally and in neighbouring countries. Range restriction is considered a better measure of conservation status than endemicity, although it is far less often used and yields very different patterns. Properly and consistently calculated measures of national endemicity do, however, retain significant conservation value, and the rates for South African marine biota are high relative to other regions globally, being exceeded only by New Zealand and Antarctica. It is important that when citing endemicity statistics, researchers and conservation managers understand the definitions used and the many constraints under which these measures are derived.

Introduction

Endemicity is a widely employed concept, used both as a proxy for range restriction, and as an indicator of the uniqueness and conservation value of the fauna and/or flora within a particular geographical (usually politically defined) region. For example, high proportions of endemicity within various South African taxa have been cited in support of the region being recognised as a 'biodiversity hotspot of international importance'.[1] Despite its wide use, endemicity is, however, a loosely defined concept – the usual dictionary definition being some variation of 'native or limited to a certain region'.

In practice, the most commonly used boundaries for 'the region' are national political borders, hence the most common regional usage is 'endemic to South Africa'. However, because authors are free to define their own boundaries, these can extend to incorporate wider politically defined regions ('southern Africa'), or a region defined by latitudinal boundaries ('Africa South of 20S') – see marine examples below. Many authors also report rates of endemism for sub-regions within the country, both for biomes, such as the 'thicket biome'[2], 'succulent karoo' or 'fynbos'[3], or for narrower regions, such as 'north-eastern Transvaal Escarpment'[4], 'southern Langeberg Mountains'[5], or 'the Cape Peninsula'[6]. In these cases, the numbers reported are of species globally unique to that biome or site (i.e. endemic to the defined region).

Another group of authors divides the region into grid squares[7,8], or the coastline into 50 km or 100 km units[9,10], then report on the numbers of endemic species present in each square or unit. It is important to note that what is then reported is numbers of species endemic to South Africa that are found in each square/unit (i.e. the same endemic species are reported for multiple squares or units). A few marine authors[10,11] do subsequently plot numbers of species unique or endemic to only one to three coastal units ('range restricted endemics'). These two methods, of course, yield very different numbers of endemics and geographical patterns of endemicity.

The marine fauna of South Africa is no exception with regard to these confusions, and many authors have published endemicity data for various taxa in the region (see Table 1), but have used a variety of definitions of endemicity, greatly complicating the interpretation of these data. The most commonly used definition, and that most comparable to data from other countries, is that of species restricted to the exclusive economic zone (EEZ) of continental South Africa. However, given the continuity of the South African EEZ with those of adjacent countries, the important role South African researchers have historically played in taxonomic research in the wider region, and other political considerations (e.g. Namibia having long been a protectorate of South Africa), several other definitions have been used. These include latitudinal definitions ('Africa south of 20S' or 'Africa south of the Tropic of Capricorn'), or wider definitions of a 'southern African' region – most often including the whole of Namibia, but only parts of Mozambique. Some studies also restricted their analyses to some maximum depth, while others included the entire EEZ. Clearly, each of these definitions results in different numerical measures of endemicity, even when based on the same data sets.

Table 1: Measures of endemicity for selected marine taxa from southern Africa. Taxa selected are 'major' grouping for which at least two separate endemicity estimates are available

Group	Reference	Region considered	Number of species	Number of endemic species	% endemic species
Mammalia	[12]	SA EEZ	43	0	0
	[13]	SA EEZ	44	0	0
Aves	[12]	SA EEZ	222	0	0
	[13]	SA EEZ	222	0	0
Reptilia	[12]	SA EEZ	6	0	0
	[13]	SA EEZ	6	0	0
Pisces	[12]	SA EEZ	2000	280	14
	[9]	SA <200m	1239	101	8
	[13]	SA EEZ	2000	280	14
Ascidiacea	[10]	SA <100m	134	72	54
	[12]	SA EEZ	140	81	58
	[14]	Namibia-central Mozambique; <15m	135	85	63
	[15]	Angola-southern Mozambique including islands	168	79	47
	[13]	SA EEZ	145	45	31
Echinodermata	[16]	Africa south of 21.5^{0}S	280	134	48
	[17]	Africa south of 21.5^{0}S	407	190	47
	[12]	SA EEZ	410	187	46
	[10]	SA; <100m	65	12	19
	[14]	Namibia-central Mozambique; <15m	150	65	43
	[13]	SA EEZ	410	187	46
Platyhelminthes	[12] (incorrectly summed as 28 with 17 endemic in that paper)	SA EEZ	56	34	61
	[13]	SA EEZ	254	17	7
Nematoda	[12]	SA EEZ	338	30	9
	[13]	SA EEZ	338	30	9
Copepoda	[12]	SA EEZ	429	41	10
	[13]	SA EEZ	429	41	10
Cirripedia	[12]	SA EEZ	86	23	27
	[13]	SA EEZ	86	23	27
	[18]	SA EEZ	86	22	25
Amphipoda	[19]	Africa south of 20^{0}S	292	134	46
	[12]	SA EEZ	454	149	33
	[10]	SA <100m	194	77	40
	[14]	Namibia-central Mozambique; <15m	189	77	41
	[13]	SA EEZ	454	149	33
Cumacea	[20]	Africa south of 20^{0}S	75	67	89
	[12]	SA EEZ	98	70	71

Group	Reference	Region considered	Number of species	Number of endemic species	% endemic species
	[13]	SA EEZ	98	70	71
Isopoda	[12]	SA EEZ	300	255	85
	[10]	SA <100m	252	224	84
	[14]	Namibia-central Mozambique <15m	278	231	83
	[13]	SA EEZ	300	255	85
Decapoda	[12]	SA EEZ	750	150	20
	[13]	SA EEZ	750	150	20
Polychaeta	[12]	SA EEZ	760	160	21
	[10]	SA <100m	523	108	21
	[14]	Namibia-central Mozambique <15m	556	170	31
	[13]	SA EEZ	760	161	21
Polyplacophora	[12]	SA EEZ	29	26	90
	[10]	SA <100m	23	18	78
	[14]	Namibia-central Mozambique <15m	24	22	92
	[13]	SA EEZ	29	26	90
Gastropoda	[12]	SA EEZ	2262	1258	56
	[13]	SA EEZ	2262	1258	56
(subgroups combined)	[10]	SA <100m	998	552	55
	[14]	Namibia-central Mozambique <15m	919	575	63
Bivalvia	[12]	SA EEZ	560	270	48
	[10]	SA <100m	208	93	45
	[14]	Namibia-central Mozambique <15m	200	89	45
	[13]	SA EEZ	650†	270	42
Bryozoa	[12]	SA EEZ	280	99	34
	[13]	SA EEZ	270	172	63
Actinaria	[12]	SA EEZ	43	19	44
	[21]	SA EEZ	47	23	49
	[13]	SA EEZ	43	19	44
	[22]	SA EEZ	61	28	46
Octocorallia	[12]	SA EEZ	204	110	54
	[10]	SA <100m	54	16	30
	[14]	Namibia-central Mozambique <15m	68	55	81
	[13]	SA EEZ	204	110	54
Hydrozoa	[12]	SA EEZ	457	86	19
	[13]	SA EEZ	382	86	23
Porifera	[12]	SA EEZ	289	10	3
	[13]	SA EEZ	346	32	9
Combined fauna	[12]	SA EEZ	11 130	3496	31
	[10]	SA; <100m	2533	931	26

Group	Reference	Region considered	Number of species	Number of endemic species	% endemic species
	[14]	Namibia-central Mozambique <15m	2650	1401	53
	[13]	SA EEZ	12 914	4233	33
	[23]	SA EEZ	12715	3549	28

[1]*Probably transposed incorrectly from 560 of Gibbons et al.*[12]

For more complete taxonomic listings, see Gibbons et al.[12] *(Table 1) and Griffiths et al.*[13] *(Table S1). Geographic region and depth zones are specified for each estimate (EEZ = Exclusive Economic Zone).*

Actual levels of endemicity are also subject to constant change over time in response to developments in taxonomic knowledge, both within and outside of the region under consideration. Simply put, new taxonomic data from within the study region will *increase* endemicity when new species are described from within the study area, but will *reduce* endemicity when species already known elsewhere are newly reported there. Conversely, studies done *outside* of the study area will depress endemicity levels if they record species previously thought to be endemic to South(ern) Africa. Keeping track of these changes requires a high level of taxonomic expertise and many regional publications have in fact based their analysis on historical data, derived from dated taxonomic monographs or museum catalogues, without correcting these to take account of more recent taxonomic revisions and discoveries, particularly those carried out *outside* of the study region. Such estimates must thus be treated with caution.

Another group of studies have analysed how the numbers, or proportions, of endemic marine species vary around a series of equal 50 km or 100 km units distributed around the length of the South(ern) African coastline. Such studies include those on fish[4] and on various invertebrate taxa.[10,11,13.] In addition to the biases and constraints already mentioned, there is also an inherent bias in such 'within region' analyses, in that such levels of endemicity are partially determined by distance away from the defined boundary. In other words, if all species had randomly-centred 500 km ranges, all species from sites more than 500 km from the defined boundary would by necessity be considered endemic. In this context, measures of range-restriction may be more indicative of rarity or conservation status than the far more commonly-used endemicity, although they too suffer from biases related to sampling effort, as discussed below.

However flawed they are in biological terms, endemicity statistics do still have conservation value, as they do give an indication of the number (or proportion) of described species that fall under the responsibility of a particular national government or conservation agency, or of the proportions that are conserved within existing protected areas[10]. Thus a critical appraisal of the use of endemicity statistics is necessary if we are to ensure that prioritisation of conservation goals is underpinned by sound scientific inputs, of which endemicity is an important example.

The South African marine environment provides a particularly good, and unusual, model system to undertake such an analysis. Firstly, a large number of regional endemicity measures have been published and (unusually in the global context) a variety of definitions have been applied to delimit the region under consideration. The South(ern) African marine environment is also a good model to analyse in that the coastline in almost linear, with very few inlets, bays or islands, plus almost all samples have been collected within a fairly narrow band adjacent to that coastline.[13] This means that it is easy to re-analyse the data in terms of linear distance from the boundary, or in terms of linear measures of range-restriction.

In this paper we aim to explore the use and usefulness of the concept of endemicity, as applied to the marine fauna of South(ern) Africa. We do this in several ways: firstly, by reviewing previous studies that have calculated endemicity and comparing the data so obtained in the light of the definitions and data sources used and, secondly by discussing factors that affect reported levels of endemicity. Lastly, we compare measured marine endemicity statistics in South Africa with those from other comparable regions and debate the usefulness of alternative measures.

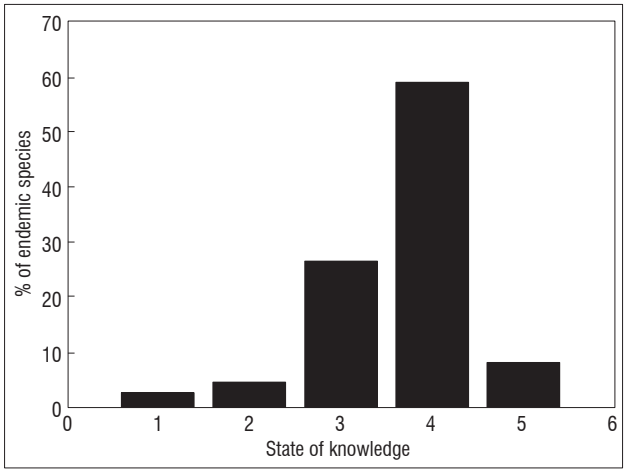
Existing measures of endemicity

Endemicity measures have been published for a large variety of South(ern) African marine taxa, of which we have reproduced only a selected subset of 'major' groups, especially focussing on those for which multiple estimates are available (Table 1). The tremendous variation in absolute measures of endemicity (from zero in several vertebrate taxa, to over 80% in Cumacea, Isopoda and Polyplacophora) is notable. There is also considerable variability between estimates made for the same taxonomic group, but by different authors. These may have resulted from inclusion of different data sets, but most obviously vary with the definition of the study area, with larger study areas generally resulting in the inclusion of the ranges of more species and thus in higher endemicity levels. For some taxa the various estimates are identical, as either the exact same data set has been analysed, or because later authors have cited figures directly from earlier ones, without any new analyses. In both cases, inaccuracies are likely to result, as in reality, new research both inside and outside the study area will have resulted in changing proportions of endemics, as discussed above.

Comparing groups, one would expect pelagic taxa to have much lower endemicity levels that benthic ones, as pelagic taxa are far more mobile. This certainly applies to some taxa, for example, the Copepoda, the majority of which are pelagic, have much lower endemicity figures than most benthic invertebrate groups. Indeed, if pelagic taxa within the Copepoda are separated from benthic ones, this trend becomes dramatic – pelagic copepods being only 1% endemic and benthic ones 49% endemic.[12] The same is true within the Amphipoda, which also contain both pelagic taxa (Hyperiidea), which are only 1% endemic, and benthic taxa (Gammaridea and Caprellidea), which together show 45% endemicity.[12] One might also expect that broadcast spawning taxa might have lower endemicity levels that brooding ones, but clear distinctions among benthic taxa, based on life history, are not apparent within these data sets. For example, while endemicity levels among some brooding taxa, such as Cumacea (71–89%) and Isopoda (83–85%) are, as expected, exceptionally high, equally high figures are evident among (mainly) broadcast spawning groups, such as Polyplacophora (78–92%) (Table 1). Conversely, Echinodermata (generally broadcast spawners) have lower endemicity ratios (19–48%), similar to the brooding Amphipoda (33–46%).

The states of taxonomic knowledge, both within the region of study and in adjoining areas, have major impacts on reported levels of endemicity. Clearly, if few species have been described from within a study region, there can be few endemic species reported from that region. Thus one would expect taxa that have been poorly studied within South(ern) Africa to have artificially depressed levels of reported endemicity. This is true, for example, for Nematoda at 9%. Conversely, for well-studied taxa, most of the actual endemic species present would have been discovered and described and hence endemicity ratios should be higher (and more realistic). These expectations are validated when plotting the relationship between numbers of endemic species and the state

of knowledge of various marine taxa for South Africa (Figure 1). Note that this relationship breaks down for the few taxa in the highest state of knowledge category, simply because these include only vertebrate groups and mangroves, which comprise large, highly mobile species with large ranges, and brown algae, for which no endemicity data are available.[13] While large size seems to equate with low endemicity in those taxa, as is also reported by Costello et al.[24], only a very weak positive correlation between size and range was observed among South African invertebrate taxa.[14]



Source: Table S1 in Griffiths et al.[13]

Figure 1: The proportions of endemic species in various taxonomic groups relative to their states of taxonomic knowledge as rated from poor (1) to excellent (5).

Additional biases in reported levels of endemicity are also created by state of knowledge in adjoining regions. Thus if the fauna, or a particular taxonomic group, is well known within the study area, but poorly know in adjoining regions, this artificially inflates the reported endemicity ratio, because species whose real ranges extend into adjoining regions are unlikely to have been detected there. Conversely, if the fauna in adjoining regions is well known, then species with broad ranges will most likely have been detected there, and endemicity within the study area will be lower (and more realistic).

An interesting example of this is the comparison between the biologically similar and closely related Amphipoda and Isopoda (Table 1). The former are well studied in Madagascar[25,26] and have a relatively low endemicity ratio, while the Isopoda, which are well known in South Africa,[27] but poorly studied elsewhere in Africa, have a much higher apparent proportion of 'endemic' species.

The effects of increased study area on the overall endemicity ratio of the fauna as a whole is particularly clear, with Gibbons et al.[12], Awad et al.[10], Costello et al.[23] and Griffiths et al.[13] giving similar overall marine endemicity estimates for South Africa (31%, 26%, 28% and 33% respectively), while Scott[14] gives a much higher estimate of 53% for the inclusive southern African region (Namibia to central Mozambique). Note, however, that even this comparison is not based on identical data sets, as each of these studies included a somewhat different number and mix of taxa, and both Awad et al.[10] and Scott[14] considered only coastal fauna, whereas Gibbons et al.[12], Costello et al.[23] and Griffiths et al.[13] examined the entire EEZ (and derived much of their data from the same sources). When we look at individual taxa, the trend remains for higher endemicity whenever a wider area is examined. For example, 55–56% endemicity for South African Gastropoda vs 63% for the wider southern African region; 21% of South African Polychaeta vs 31–36% for the wider region; 71% vs 89% for Cumacea, etc. There are, however, exceptions, such as Primo & Vazque[15], who give a lower endemicity estimate of 47% for the Ascideacea of the wider Angola–Mozambique region than either Gibbons et al.[12] or Awad et al.[10] (but not Griffiths et al.)[13] provide for South Africa alone (58%, 54% and 31% respectively). This can only be the result of different data sets being used in these analyses.

Several studies have reported on geographical patterns of endemicity within the region,[9,10,11] plotting patterns of both species richness and endemicity within particular taxa as one moves around the coastline of South(ern) Africa (Figure 2). The usual findings are that numbers (and proportions) of endemic species peak along the South Coast. There are good biological reasons to expect this trend, in that the warm temperate South Coast is both situated exclusively within the political boundaries of South Africa and is geographically isolated from other warm temperate regions (both features which promote high endemicity). By contrast, the cool temperate west coast and more tropical east coast bioregions are contiguous with similar climatic regions in adjacent countries, meaning that species living in those habitats are likely to range across political borders and hence not be endemic.

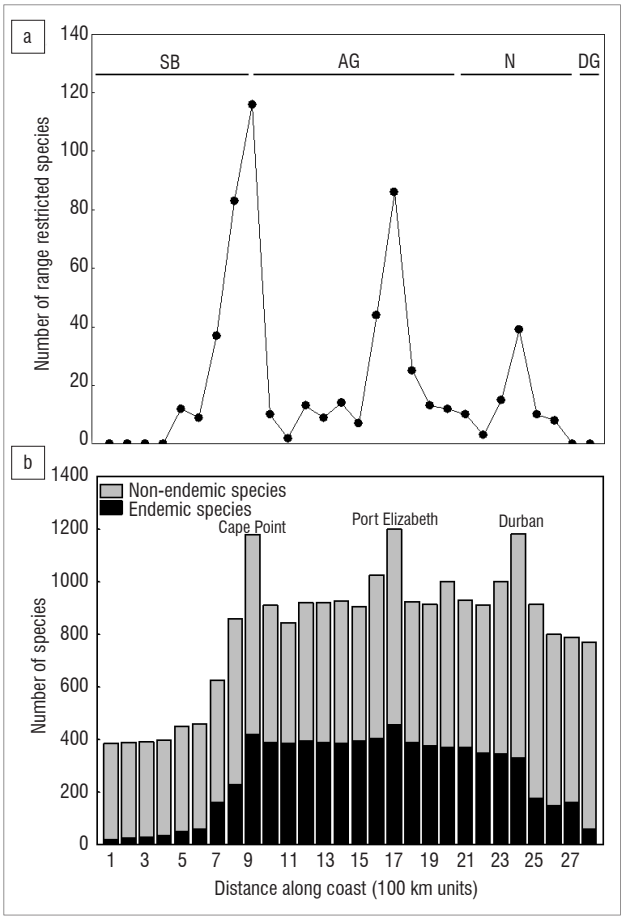


Figure 2: (a) Numbers of range-restricted (range<300 km) species/100 km stretch of coastline and (b) marine invertebrate species richness/100 km. Distance around coast measured from Namibian border to Mozambique border. Data from Awad et al.[10] Ecoregions described by Sink et al.[28] are indicated in panel (a) as SB (Southern Benguela), AG (Agulhas), N (Natal) and DG (Delagoa). The dark sections of the bars in panel (b) indicate endemic species.

It is important to note here, though, that endemicity is strongly influenced simply by distance from the political borders of South Africa and is not a good indicator of either range restriction or rarity. For example, a species with a short range that happens to straddle a political border will not be endemic to either country, but may have a very restricted range and hence require conservation attention. Conversely, one that is endemic can have a much larger range (of up to 3000 km in this case) and hence be of less conservation concern. Endemicity along the South Coast of South Africa is thus inflated simply because this region is furthest from the political borders used to define endemicity in this region.

Marine endemicity rates in South Africa in the global context

Table 2 shows known rates of marine endemicity in a variety of global regions, as summarised by Costello et al.[23] It is notable that South Africa supports the third highest proportion of endemic species among listed regions, despite having one of the smaller seabed areas and supporting comparatively few described species, relative to some other listed regions. The reasons for these pattern are complex and involve not only 'real' factors, such as the size, isolation and taxonomic uniqueness of the fauna of regions under consideration, but artefacts associated with sampling and taxonomic effort, both in the region in question and in neighbouring areas.

Table 2: Numbers of marine species reported from various regions globally, the percentage of these that are endemic and size of region.

Region	Total number of species	% endemics	Seabed area km²
Antarctica	8200	45	21 186 153
Australia	32889	28	6 819 501
Baltic	5865	2	411 218
Caribbean	12046	13	2 828 125
China	22365	7	831 966
Japan	32777	6	3 970 743
Mediterranean	16845	7	2 451 059
New Zealand	12780	51	4 073 895
South Africa	12715	28	846 463

Source: Costello et al.[23]

Regions that are large and/or isolated (Australia, New Zealand, Antarctica) would be expected to, and do, have high levels of endemicity. Conversely, those with relatively small areas and/or which lie adjacent to other well-studied areas (China, Japan, Baltic, Mediterranean, Caribbean) have relatively low endemicity. South Africa might appear to be an outlier here, as it has a relatively small sea area, but a high endemicity ratio. However, it does incorporate an 'island' of warm temperate water that is very isolated geographically, from other warm temperate regions, and which hosts the majority of the endemic species. It is also something of an 'island of taxonomic expertise' in a much wider African region, within which marine biodiversity is poorly documented, and this tends to (artificially) elevate rates of endemicity.

Is range restriction a better measure than endemicity?

The problems outlined above, with respect to accurate determination of endemicity rates, beg the question whether endemicity is a useful concept for managers. Clearly, it does have political importance in that the conservation of endemic species falls exclusively under the jurisdiction of a particular nation's conservation authorities. It is also a term the public associate with and which has implications for funding of conservation efforts through eco-tourism, etc.

On the other hand, range restriction and population size or rarity (which are not dealt with here, are more difficult to measure and more rarely reported in marine taxa) are much more important concepts in terms of viability and vulnerability of the population of a given species. Studies by Awad et al.[10] and Scott et al.[11] have plotted the geographic distributions of range-restricted species along the South Africa coast, showing range-restricted forms to be strongly localised relative to endemic ones. Specifically, range-restricted marine species tend to be strongly concentrated at focal sites in the vicinities of Cape Point, Port Elizabeth and to a lesser extent, Durban (Figure 2a). Scott et al.[11] suggest that, as these sites tend to correspond with biogeographic breaks, these peaks may represent species living at the ecotone between biogeographic provinces. However, these specific locations coincidentally also fall at the locations of the three major coastal centres of marine research in the country and are, therefore, also sites of enhanced sampling effort and species discovery. As these well-sampled areas thus contain 'extra' species not reported in adjacent sectors of coast (Figure 2), it is it thus inevitable that these 'extra' species will appear to be range-restricted to those sites. Moreover, while the intensely-sampled area around Cape Point does indeed coincide exactly with the boundary between the Southern Benguela and Agulhas ecoregions,[28] the peaks around Port Elizabeth and Durban are offset from biogeographic breaks by at least 300 km. This strongly suggests that at least these peaks in locally endemic species are largely a function of differences in sampling effort, rather than true range restriction. The larger peak around Cape Point may be more 'real' as it is situated in a region with unique ecotonal characteristics, but the greatly enhanced sampling effort at this site is certainly responsible for at least some of these apparently 'localised' species.

Acknowledgements

Financial support for this research was provided by a National Research Foundation grant to CLG, though the SeaKeys programme.

Authors' Contributions

The paper was a joint effort by C.L.G. and T.B.R. C.L.G. conceived the idea for the paper and wrote the original draft. T.B.R. edited and added to the draft and drew the figures. Being a review, there were no original experiments, only data extraction and interpretation which we did together.

References

1. Endangered Wildlife Trust. The Biodiversity of South Africa 2002: Indicators, trends and human impacts. Cape Town: Struik Publishers; 2002.
2. Vlok JHJ, Euston-Brown DIW, Cowling RM. Acocks' Valley bushveld 50 years on: New perspectives on the delimitation, characterisation and origin of subtropical thicket vegetation. S Afr J Bot. 2003;69:27–51. http://dx.doi.org/10.1016/S0254-6299(15)30358-6
3. Verboom GA, Archibald JK, Bakker FT, Bellstedt DU, Conrad F, Dreyer LL, et al. Origin and diversification of the Greater Cape flora: Ancient species repository, hot-bed of recent radiation, or both? Molec Phylogenetics Evol. 2009;51:144–53. http://dx.doi.org/10.1016/j.ympev.2008.01.037
4. Matthews WS, Van Wyk AE, Bredenkamp GJ. Endemic flora of the north-eastern Transvaal escarpment, South Africa. Biol Conserv. 1993;63:83–94. http://dx.doi.org/10.1016/0006-3207(93)90077-E
5. McDonald DJ, Cowling RM. Towards a profile of an endemic mountain fynbos flora: Implications for conservation. Biol Conserv. 1995;72:1–12. http://dx.doi.org/10.1016/0006-3207(94)00030-T
6. Picker, MD, Samways MJ. Faunal diversity and endemicity of the Cape Peninsula, South Africa – a first assessment. Biodiver Conserv. 1996;5:591–606. http://dx.doi.org/10.1007/BF00137611
7. Skelton PH, Cambray JA, Lombard AT, Benn GA. Patterns of distribution and conservation of freshwater fishes in South Africa. S Afr J Zool. 1995;30:71–81. http://dx.doi.org/10.1080/02541858.1995.11448375
8. Drinkrow DR, Cherry MI. Anuran distribution, diversity and conservation in South Africa, Lesotho and Swaziland. S Afr J Zoo. 1995;30:82–90. http://dx.doi.org/10.1080/02541858.1995.11448376
9. Turpie JK, Beckley LE, Katua SM. Biogeography and the selection of priority areas for conservation of South African coastal fish. Biol Conserv. 2000;92:59–72. http://dx.doi.org/10.1016/S0006-3207(99)00063-4
10. Awad AA, Griffiths CL, Turpie JK. Distribution of South African marine invertebrates applied to the selection of priority conservation areas. Divers Distrib. 2002;8:129–145. http://dx.doi.org/10.1046/j.1472-4642.2002.00132.x
11. Scott RJ, Griffiths CL, Robinson TB. Patterns of endemicity and range-restriction among southern African coastal marine invertebrates. Afr J Mar Sci. 2012;34:341–348. http://dx.doi.org/10.2989/1814232X.2012.725284
12. Gibbons MJ. The taxonomic richness of South Africa's marine fauna - a crisis at hand. S Afr J Sci. 1999;95:8–12.

13. Griffiths CL, Robinson TB, Lange L, Mead A. Marine biodiversity in South Africa – state of knowledge, spatial patterns and threats. PLoS ONE 2010;5(8):e123008. http://dx.doi.org/10.1371/journal.pone.0012008

14. Scott R. 2009. Biogeographical patterns in southern African marine invertebrates [MSc dissertation]. Cape Town: University of Cape Town; 2009.

15. Primo C, Vázquez E. Zoogeography of the southern African ascidian fauna. J Biogeogr. 2004;31:1987–2009. http://dx.doi.org/10.1111/j.1365-2699.2004.01144.x

16. Clark AM, Courtman-Stock J. The Echinoderms of Southern Africa. London: British Museum (Natural History); 1976.

17. Thander AS. Zoogeography of the southern African echinoderm fauna. S Afr J Zoo. 1989;24:311–318.

18. Biccard A. Taxonomy, systematics and biogeography of South African Cirrepedia (Thoracica) [MSc dissertation]. Cape Town: University of Cape Town; 2012.

19. Griffiths CL. The gammaridean and caprellid Amphipoda of southern Africa [PhD thesis]. Cape Town: University of Cape Town; 1974.

20. Day JA. Southern African Cumacea [PhD thesis]. Cape Town: University of Cape Town; 1978.

21. Acuña FH, Griffiths CL. Species richness, endemicity and distribution patterns of South African sea anemones (Cnidaria: Actiniaria & Corallimorpharia). Afr Zool. 2004;39(2):193–200.

22. Laird CM. Taxonomy, systematics and biogeography of South African Actiniaria and Corallimorpharia [PhD thesis]. Cape Town: University of Cape Town; 2013.

23. Costello MJ, Coll M, Danovaro R, Halpin P, Ojaveer H, Miloslavich P. A census of marine biodiversity knowledge, resources and future challenges. PLoS ONE 2010;5(8):e12110. http://dx.doi.org/10.1371/journal.pone.0012110

24. Costello MJ, May RM, Stork NE. Can we name Earth's species before they go extinct? Science. 2013;339(6118):413–416. http://dx.doi.org/10.1126/science.1230318

25. Ledoyer M. Crustacés amphipodes gammariens. Familles des Acanthonotozomatidae à Gammaridae [Gammaridean amphipod crustaceans. Families Acanthonotozomatidae to Gammaridae]. Faune de Madagascar. 1982;59:1–598. French.

26. Ledoyer M. Crustacés amphipodes gammariens. Familles des Haustoriidae à Vitjazianidae [Gammaridean amphipod crustaceans. Families Haustoriidae to Vitjazianidae]. Faune de Madagascar. 1986;59:599–1112. French.

27. Kensley B. Guide to the marine isopods of Southern Africa. Cape Town: Trustees of the South African Museum; 1978.

28. Sink K, Holness S, Harris L, Majiedt P, Atkinson L, Robinson T, et al. National Biodiversity Assessment 2011: Technical Report. Volume 4: Marine and Coastal Component. Pretoria: South African National Biodiversity Institute; 2012.

Enhancing climate governance through indigenous knowledge: Case in sustainability science

AUTHORS:
Nelson Chanza[1]
Anton de Wit[2]

AFFILIATIONS:
[1]Department of Geography, Bindura University of Science Education, Bindura, Zimbabwe
[2]Department of Geosciences, Nelson Mandela Metropolitan University, South Africa

CORRESPONDENCE TO:
Nelson Chanza

EMAIL:
nchanza@gmail.com

POSTAL ADDRESS:
Department of Geosciences, Nelson Mandela Metropolitan University, PO Box 77000, Port Elizabeth 6031, South Africa

DATES:
Received: 18 Aug. 2014
Revised: 06 May 2015
Accepted: 10 Oct. 2015

KEYWORDS:
climate change; local participation; disaster management; EBA; mitigation; CBA

HOW TO CITE:
Chanza N, De Wit A. Enhancing climate governance through indigenous knowledge: Case in sustainability science. S Afr J Sci. 2016;112(3/4), Art. #2014-0286, 7 pages. http://dx.doi.org/10.17159/sajs.2016/20140286

The current tempo of climate change strategies puts the notion of sustainability in question. In this philosophy, mitigation and adaptation strategies ought to be appropriate to the sectors and communities that are targeted. There is a growing realisation that the effectiveness of both strategies hinges on climate governance, which also informs their sustainability. The application of the climate governance concept by the technocratic divide (policymakers and climate practitioners) to communities facing climate change impacts, however, is still a poorly developed field, despite extensive treatment by academia. By drawing heavily from conceptual and analytical review of scholarship on the utility of indigenous knowledge (IK) in climate science, these authors argue that IK can be deployed in the practice of climate governance. It reveals that the merits of such a deployment lie in the understanding that the tenets of IK and climate governance overlap and are complementary. This is exhibited by examining the conceptual, empirical and sustainability strands of the climate governance-IK nexus. In the milieu of climate change problems, it is argued that the basic elements of climate governance, where actions are informed by the principles of decentralisation and autonomy; accountability and transparency; responsiveness and flexibility; and participation and inclusion, can be pragmatic particularly to communities who have been religiously observing changes in their environment. Therefore, it becomes necessary to invigorate the participation of communities, with their IK, in designing climate change interventions, which in this view can be a means to attain the objectives of climate governance.

Introduction

The Intergovernmental Panel on Climate Change (IPCC), a brainchild of the United Nations Environmental Programme (UNEP) and the World Metrological Organisation (WMO) established to give the most comprehensive overview and fact base of climate science, recognises climate governance and indigenous knowledge (IK) in the ongoing climate change discourse. Both concepts, however, are treated in fragmentation. Despite this segmentation, there is a growing appetite by both climate governance community and indigenous knowledge researchers to examine the usefulness of the concepts in climate interventions. In climate regimes, it can be argued that climate governance has dominated climate discussions more than IK.[1-3] This dominance continues to grow against a backdrop where the intergovernmental policy arena faces substantial impasse on what constitutes climate governance. The intricacy of the stalemate arises from a tradition characterised by a disparate magnitude of contribution existing between the global North, largely blamed for greenhouse gas (GHG) emissions, and the global South, paradoxically experiencing extreme climatic events. Indigenous climate knowledge, following at a distance behind climate governance, continues to occupy space in climate literature, albeit with limited realisation of the relationship between these concepts. It is therefore essential to advance discourse by examining the relationship between climate governance and IK. In this paper, we argue that IK has potential to transform the technocratic-community engagement front in the current discussions where climate science and policy regimes are increasingly being interrogated for their sustainability.

Conscious of the conceptual and definitional flaws, and the intricacies characterising climate discourse, we proceed by treating climate governance as a concept embracing inclusivity in designing mitigation and adaptation strategies by all climate stakeholders, including indigenous communities affected by climate change. Adaptation governance, a concept that has dominated climate discussions to date, centres, for example, on participation, equality and justice in decision-making about interventions to contain climatic events. Meadowcroft[1] says that, at country level, this requires adequate knowledge about anticipated climate effects and planning to tackle anticipated impacts on human activity. On the other hand, mitigation governance calls for an understanding of emission sources, cost-effective containment potentials, and policy approaches.[1,2]

In a related treatment, IK is a knowledge form defined by Orlove et al.[4] as place-based and rooted in local cultures that are mostly associated with long-settled communities who have strong ties to their natural environments. The concept is now increasingly acknowledged by the IPCC as evidenced by distinct sections covering IK in its latest reports.[5-7] The disaster management community is also drawing heavily from the experiences of indigenous people with the view of making impacting risk and disaster management interventions.[8-12]

Evidence of IK's usefulness in climate science ranges from enhancing understanding of climate impacts,[13-15] particularly at local scale where scientifically advanced models tend to give a coarse-grained focus,[16] to informing successful mitigation and adaptation interventions[13,17-19] whose success could be credited on meaningful community participation in identifying appropriate climate projects.[15,20-22] A demonstration of the climate governance-IK linkages is given in this paper. This is exclusively done through reviewing literature on climate governance and indigenous climate knowledge across the world.

The overriding theme in this analysis is the interrogation of the notion of sustainability science. The genesis of this philosophy is traced from the United Nations Educational, Scientific and Cultural Organisation (UNESCO) Declaration on Science and the Science Agenda Framework held in Budapest, Hungary in 1999.[23] Earlier,

sustainability as a concept gathered impetus following the 1992 World Earth Summit on Environment and Development held in Rio de Janeiro, Brazil[24] but was preceded by the 1987 Brundtland Report[25] that specified a framework for sustainable development articulation. Within this view, in order for scientific interventions to be sustainable, a platform that incorporates the active involvement of citizens who should be served by the science ought to exist. In the context of climate governance therefore, climate science should be seen by local people as a shared asset that helps them to seek practical solutions to the problems and opportunities brought by change and variability in the climate system. The paper implores that the praxis of climate governance at local level is potentially realisable when indigenous communities, with their invaluable reservoirs of climate change knowledge, actively participate in climate regimes. Four cases relevant in climate mitigation and adaptation are thus drawn here:

- The envisaged benefits of local participation in programmes like that on Reducing Emissions from Deforestation and Forest Degradation (REDD+), a mechanism whose thrust is to enable communities in developing countries to benefit from climate funds if they actively partner in forest projects that enhance carbon sinks
- Land Use, Land-Use Changes and Forestry (LULUCF), intended for atmospheric carbon stabilisation through regulated activities in local land-use planning and management
- Community-based adaptation (CBA), a bottom-up framework for making effective adaptation through the central role of local people
- Ecosystem-based adaptation (EBA), an approach that serves both objectives of climate proofing and emission reduction through harnessing ecosystem services coupled with wise management of the same.

The benefits of a climate governance-indigenous knowledge nexus are discussed here from conceptual significance, empirical evidence to sustainable development. But first, it is crucial to trace the major developments on governance around climate change so as to contextualise the discussion.

Nature of climate governance and problems of articulation

There is phenomenal growth in literature that examines the concept of climate governance. However, this has not resolved the current impasse in the application of the concept in the global policy arena that has been in existence for the past two decades. Biermann and Boas[26] argue that the difficulties emerge from the intricacies of the phenomenon itself, where climate change is seen as a problematique both in causation and consequences, where the industrialised societies, largely to blame for the anthropogenic forcing of climate destabilisation, should have a moral responsibility to contain its devastating impacts. Most authorities agree that complexities in conceptualising climate governance mirror both the multi-layered spatial scale and multi-sectoral levels of application.[27-32] Van Asselt,[31] for example, views this as a fragmentation of climate governance. Bulkeley and Moser[33] agree with Andonova et al.[28] that climate governance needs to be decentralised beyond multilateral agreements, and diffused across many actors in society. Backstrand and Loubrand[29] pose that the concept has to be understood from competing discourses of green governmentality, ecological modernisation and civic environmentalism, where '…local is pitted against global, North vs South, public vs private, decentralisation vs centralisation, and economic efficiency vs environmental integrity.' Other scholars, like Pattberg and Stripple[30] and Saran[34], adopt the term global governance, arguing that the challenges of climate change are global in dimension and cannot be addressed by national or regional interventions alone. Clearly, it can be inferred from these authorities that climate governance is a multifaceted term with a varied scale of application.

At a higher level, governance of climate change is predominantly seen in the 1992 United Nations Framework Convention on Climate Change (UNFCCC) and the 1997 Kyoto Protocol. The UNFCCC is designed to stabilise GHG concentrations to levels that would not interfere with the earth's climate system, while in addition to sharing the objectives of the Convention, the Kyoto Protocol commits industrialised countries to stabilise their GHG emissions. Ashton and Wang[27] are of the view that the two regimes reflect a general calculus of equity through the principle of common but differentiated responsibilities. Under this framework, attention can be given to poor and vulnerable societies to acquire technology and funds to adapt to climate change. However, the intergovernmental policy landscape still faces challenges in articulating the idiosyncrasies of climate impacts permeating diverse sectors that are not clearly specified in the multilateral climate agreements. Van Asselt[31] argues that the international environmental law in place has not been able to comprehensively deal with the issue of global climate governance. One of the areas not clarified by these regimes is the phenomenal increase in the population of climate refugees as noted by Biermann and Boas[26] and Martin[35].

At the regional level, for example, the European Union (EU) and Southern African Development Community (SADC), climate governance can be effected through appropriate regional environmental policies and protocols. Oberthür[36] is of the view that the limitations of the current global climate institutional architecture can be reinforced through strengthening regional climate policies, such as that of the EU. In Africa, for example, these regimes have been characterised by a fragmentation of responses, which has not done justice to the demands set by climate governance.[28]

Nationally, climate governance is tackled in the form of climate policies and institutions. In most countries, however, the policy domain is marked by a patchwork of public and private institutions that differ in their mandates and interests, with potential to cause conflicts that are likely to stymie governance issues.[26] Meadowcroft's[1] paper looks at national level climate governance, which can be influenced by promoting coalitions for change, minimising antagonistic forces, establishing centres of economic efficiency, creating robust institutions, adjusting national environmental legislations, and transforming citizen behaviour. He cites the problem of institutional inertia, where perceptions about adverse consequences of mitigation policies in combination with scientific uncertainty tend to block progress towards the attainment of global climate governance.

In order to address some of these problems, United Nations Development Programme/ United Nations Capital Development Fund/ United Nations Environment Programme (UNDP/UNCDF/UNEP)[2] identifies 5 entry points for enhancing the way in which national or local governments can interface with climate change:

1. National climate change and sub-national governance policies should involve clear links between national policies and decentralisation/ sub-national policies, especially in developing countries where such linkages are best described as weak or non-existent
2. Improving local understanding of climate change issues; that is, knowledge of what climate change means to local governments in concrete and tangible terms, such as information on nature and risks they face
3. Financing arrangements for climate change; including corruption proofing, where there is room for information transparency and oversight on resource use by community stakeholders
4. Operationalising local democracy for vulnerable groups such as involving indigenous peoples in climate change decision-making processes and adaptation strategies, which would include the use of indigenous knowledge and innovations
5. Addressing the capacity question, where capacity assessment of institutions, organisations and individuals is interrogated

Because of these problems of articulation at differentiated spatial and sectoral levels, the emerging situation is that climate governance remains a rhetorical commitment, particularly when viewed from the perspectives of poor and vulnerable communities relying on climate-sensitive liveli-

hoods. Despite increasing global certainty about anthropogenic forcing of climate change and the challenges it poses on humanity noted by the IPCC's[5] report, Oberthür's[36] remarks resonate with Pattberg and Stripple's[30] earlier observation that the policy architecture in place is incapable of effectively addressing climate change. Here we propose a configuration that links climate governance and IK in order to operationalise climate governance, with the intention of offering directions for building resilient sustainable communities in many parts of the world experiencing climatic events.

Climate governance-IK linkage: A conceptual necessity

Governance refers to the ways in which scale level decision-making takes place. It is called 'good governance' when it reflects scrutinisation and oversight by citizens, openness and transparency and participation.[2,32] The IPCC[7] views governance in climate change as a more inclusive approach that recognises the various levels of government (global, international, regional, national, local) and the responsibilities of the private sector, non-governmental actors and the civil society. The problem of conceptualisation has been widely reported by other scholars such as Chirisa and Chanza[32], Bulkeley and Moser[33], Pattberg and Stripple[30,37], Okereke and Bulkeley[38], and Macey[39]. Macey[39] even suggests a redesign of the climate governance framework from scratch, while Bulkeley and Moser[33] recommend a shift from conceptual necessity to empirical reality. Chirisa and Chanza[32] adopt a local governance perspective in order to demystify the institutional strictures, dispel corrupt tendencies and empower citizens for participatory democracy.

Apparently, most scholarly work is silent regarding examining the role of indigenous knowledge in climate governance. For example, scholars like Orlove[4], Nyong et al.[13], Berkes[18], Mawere[40,41], and Mawere et al.[42] who have written extensively on IK applications in climate change, have not embraced this concept. Espousing the spatial dimensions of applicability featuring prominently in many scholarly work[1-3,28,33] and policy regimes,[30,36] a re-examination of climate governance would point towards climate governance-indigenous knowledge connectedness. This linkage is argued here as a conceptual necessity for premising climate governance.

IK as a concept, defines knowledge that is location specific, acquired in situ, through progressive study of the community's interaction with the environment, and orally transferred both within and between generations.[4,40-52] Mawere[40,41] sees IK as useful in establishing a moral virtuous society whose usage helps African communities to realise a sense of responsibility in environmental resource exploitation, thus ensuring food security in the event of environmental shocks like drought. To Shizha[43], this knowledge strengthens the education of the African population, which is a necessary foundation for sustainable development. Some key issues related to governance can be inferred from this conceptualisation. Firstly, the knowledge is understood collectively to include community skills, technologies and practices that give the community collective understanding and responsibility to sustainably utilise the environment. Secondly, in both theory and praxis, the concept is holistic, empowering and participatory. As can be seen in Figure 1, IK elements are a function of climate governance. Here the major elements characterising climate governance are also found in the utilisation of IK. For instance, the knowledge form is decentralised at grassroots level structures in an autonomous arrangement where there is no exclusion in accessing the knowledge. Whenever a response to climatic events such as floods and drought is called for, users of IK are flexible to actively choose available coping and adaptive options for climate proofing. Although not clearly referring to the concept of climate governance in their arguments, Gwimbi[10], Mawere et al.[42] and Chanza[44,45] write about the significance of IK in mitigating the negative effects of climate change such as floods, violent storms, dry spells and drought.

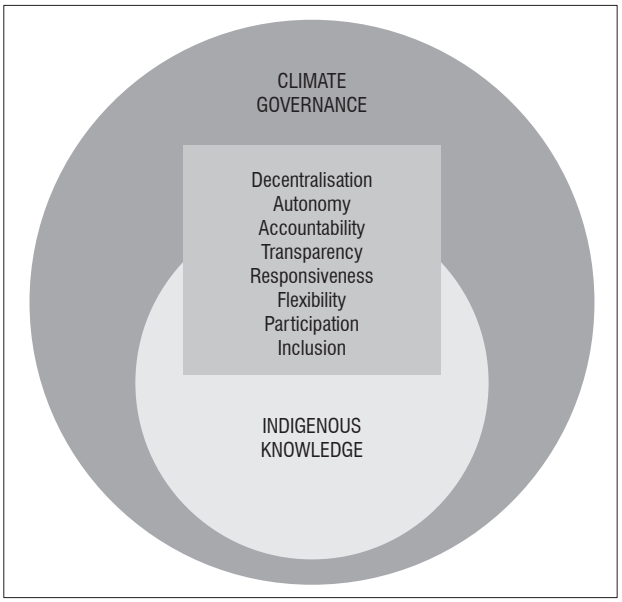


Figure 1: Basic elements of climate governance and indigenous knowledge.

Climate governance-IK linkage: An empirical reality

The climate change community cites mitigation and adaptation as central themes governing climate interventions. Mitigation, in climate science deployment, is a dual-faceted concept: firstly, as a human intervention to reduce sources or enhance sinks of GHGs; and secondly, as the moderation of potential adverse impacts of climate risks and disasters through actions that reduce hazard, exposure and vulnerability.[6] In order for local mitigation programmes to be effective, they need the support of key stakeholders that include local citizens where activities are implemented. Two interventions used by the IPCC and the UNFCCC worth noting here are REDD+ and LULUCF.

REDD+ is a carbon compensation programme that covers the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries.[53] In the planning and implementation process of the programme, the full and effective participation of indigenous people who use their IK to manage forestry resources can be realised. Through traditional resource management regimes, indigenous people can translate climate governance into practice. For example, the effectiveness of REDD+ projects in Brazil, Indonesia, Kenya, Mexico, Peru and Tanzania is attributed to participatory planning involving local citizens whose knowledge is crucial in the technical analysis to address the drivers of deforestation and forest degradation, as well as barriers to sustainable management.[54]

The fifth assessment report (AR5) of the IPCC reports that GHG emissions from agriculture comprised about 12% of manufactured pollutants in 2010.[7] This makes LULUCF a critical area for mitigation strategies. Furthermore, the Kyoto Protocol identifies changes in carbon stocks and GHG emissions by sources and removals by sinks related to direct anthropogenic land-use change and forestry activities.[55] In order to operationalise this strategy at local level, customary laws and practices governing land-use, land and forestry management activities in rural communities of most African societies, for example, can be drawn upon. In Zimbabwe, for instance, Mahamed-Katerere[56] concludes that land-use practices are interwoven with cultural beliefs under the administration of local traditional leaders in a philosophy she calls 'environment-spiritual connection'. Within this belief system, she cites two laws that guide sustainable land management. The first is a collection of rules that link abuse of resources to spiritual sanctions. Alongside this, there are also spiritual rules that restrict use and condemn unsustainable exploitation.

In a similar case, Turner and Clifton[57] in their study targeting indigenous citizens of Hartley Bay in British Columbia, Canada, observe that the motive behind sustainable land and forestry resource utilisation is the strong indigenous belief that the resources belong to ancestry and the spirits. As such, the locals have a moral and spiritual obligation not to violate the environment. This means issues of governance are properly understood and practised at grassroots levels that shape human-environment interactions.

In climate literature, adaptation is treated as a response interventions to some climate stimuli so as to minimise harm or enhance the benefits brought about by climatic conditions or events.[5,58-60] Strategies to contain climatic events are usually practical at local or project level involving interactions with the people. In this way, the participation of indigenous people using their knowledge, skills and experiences drawn from many years of coping and adapting to changing and variable environments deserve emphasis.

One of the predominantly preferred strategies by the climate change scholars and practitioners is community-based adaptation (CBA). CBA is viewed as a community-centric approach whereby the locals are positioned as the main stakeholders in the implementation of coping and adaptive interventions.[61-63] The intention is to build adaptive capacity and enhance community resilience against the disturbances potentiated by destabilisation in the climate system. In application, it can be seen that this strategy embraces a bottom-up approach where indigenous people are given respect and space to select from a toolbox of their own adaptation options. Kirkland[63] emphasises that CBA projects are more effective if they actively solicit the input and participation of local people, which from the viewpoint of Ensor[61] is governance for community-based adaptation.

In a study to understand the value of IK in climate adaptation in the African Sahel, Nyong et al.[13] mention 5 reasons meriting giving attention to local knowledge in order to guarantee successful adaptation. These are IK:

- is rich in cultural context
- is an appropriate and sophisticated knowledge form
- increases community buy-in
- promotes equity, efficiency and environmental integrity
- leads to increased communication and understanding

Experiences by Practical Action, working with indigenous rural communities in Bangladesh, Kenya, Peru and Zimbabwe show that for adaptation governance to be realisable it ought to be '…understood as a process, through which communities gain access to skills, resources and information so that they can continuously shape their lives and livelihoods as the environment changes around them.'[61]

Another key strategy capable of operationalising the climate governance-IK linkage to advance both the goals of mitigation and adaptation governance, is that of ecosystem-based adaptation (EBA). Defined by Munang et al.,[64] EBA 'is the use of natural capital by people to adapt to climate change impacts, which can also have multiple co-benefits for mitigation, protection of livelihoods and poverty alleviation.' In other words, the concept provides multiple benefits for society and the environment as it contributes to reducing vulnerability and increasing resilience to both climate and non-climate risks.[64-67] Worth noting is that the main activities of EBA, listed by the UNFCCC[67] as vulnerability assessment, capacity building, designing policy measures and implementation, can involve local people leveraging their IK to ensure adequate climate proofing and emission reduction. The Sweden case involving ecosystem-based measures by the local farmers is given as an example (see Box 1).

Box 1: Farmers network using ecosystem-based measures to cope with uncertain climatic conditions

Adjusting management practices, including adopting traditional farming techniques, can help to increase resilience and reduce vulnerability to the effects of climate change. The east-central area of Sweden presents difficult climatic conditions for small-scale farmers who experience long winters and frequent periods of drought. This climatic uncertainty, combined with threats from pests and disease, presents challenges for sustaining livelihoods, with climate change expected to exacerbate these conditions. To build resilience, farmers in the Roslagen region began incorporating a range of ecosystem-based practices. Measures included diversification of crops in time and space to reduce the risk of crop failure, using multiple crop varieties to increase genetic diversity and pest resistance, incorporating crop rotation to revitalise soils and prevent pest infestations without reliance on chemical fertilisers and pesticides, and planting shade trees and cover crops to enhance seedling survival to cope with drought. In addition, by establishing an informal local network, the farmers were able to share best practice and local ecological knowledge. The ecosystem-based measures led to the farmers producing high-quality and organic products, whilst increasing their resilience to climate variability and change. Biodiversity and economic security has also been enhanced.

Source: UNFCCC[67]

It should be noted here that focussing on IK does not guarantee the project will be equitable, just and successful, as expected in climate governance. Saran[34] states that given the appropriate conditions, local knowledge is capable of making interventions more effective. One of the factors that guarantees successful mitigation and adaptation for local based projects is the existence of social capital,[62,68] defined by Woolcock and Narayan[69] as the standards and networks that drive people toward collective responsibility and action. Under this concept, the role of social networks, community linkages and institutional structures can be exploited to enhance the objectives of climate governance or to build community resilience against the devastating impacts of climate change.

Climate governance-IK linkage: A drive towards sustainability

In the preceding sections, we have argued that the realisation of social, economic and environmental sustainability can be enhanced by mitigation and adaptation options that are appropriately crafted through issues of participation, consultation, inclusivity, efficiency, accountability and decentralisation. The IPCCC's[6] emphasis that 'sustainability in the context of climate change is addressing the underlying causes of vulnerability, including the structural inequalities that create and sustain poverty and constrain access to resource' is worth noting. In other words, climate governance should be the means to sustainability. Backstrand and Loubrand[29] suggest that the composition of future global climate governance should address elements of fairness, burden-sharing, poverty alleviation, participatory democracy and sustainable development. The Climate Action Network[70] gives guidelines that can be followed for adaptation governance to be sustainable. The principles for adaptation governance are summarised as[70]:

- Prioritise the adaptation needs of, and ensure that resources reach, the most vulnerable, including marginalised groups, women and children, indigenous peoples, local communities and those disproportionately impacted, as well as vulnerable ecosystems, through enhancing adaptive capacity and reducing vulnerability.
- Recognise that responses will have to be based on local assessment of risks, needs and circumstances and be relevant to local people and communities.
- Maximise national, sub national and community level ownership over adaptation planning and implementation processes, and disbursement of adaptation finance, in order to enable and encourage participatory local level planning and implementation.
- Plan and implement adaptation actions in a transparent and well-documented way that is open to public scrutiny and discourse. Ensure

the representation of key stakeholders, especially representatives of vulnerable communities, marginalised groups, women, and indigenous peoples at every stage of the process as appropriate – including in the governance and disbursement of adaptation finance, planning, implementation, monitoring and reporting.

- Adopt a process-driven learning-by-doing approach on adaptation planning and implementation, respecting the Precautionary Principle while recognising the urgency to adapt in the absence of complete information and the need to develop and implement flexible plans and programmes that can be updated on the basis of new information and learning.

As the global community drifts towards the post 2015 sustainable development goals (SDGs) marking a paradigmatic shift from the millennium development goals (MDGs), there is need to retrospectively centralise discourse on the sustainability of climate interventions as climate change will remain a cornerstone of the SDGs. The issue of translating climate governance rhetoric into practice should be the central theme of such a transformation. It should be learnt here that local community knowledge, skills and experiences are key in shaping sustainable climate interventions.

Conclusion

In the milieu of climate governance rhetoric characterising the climate regime terrain, the potential role that IK can play in transforming theory into practice deserves attention. The view advanced in this debate is that of taking governance issues to the people who are facing climate impacts. It is hoped that the first point of entry for meaningful climate mitigation and adaptation is the local level where knowledge and experiences of those witnessing the climate phenomena can be harnessed. The effectiveness of REDD+ and LULUCF projects, which are being considered as driving climate stabilisation agendas in many parts of the world, should not be devoid of elements of local governance reflecting local input through IK. Similarly, if adaptation is to build community resilience and adaptive capacity, adaptation governance informed by IK and experiences of locals cannot be ignored. Prioritising local climate governance would translate into local sustainable development which could lead to potential ramifications in higher levels and towards global sustainability. Therefore, it serves to mention that climate governance dialogue should be supported starting at grassroots where indigenous people, with their banks of IK, should characterise the policy discourse for sustainability.

Authors' contributions

N.C. was the lead author who structured the paper concept. AdW reviewed the draft paper and made various suggestions to improve its readability for the target audience.

References

1. Meadowcroft J. Climate change governance: A paper contributing to the 2010 World Bank World Development Report. Policy Research Working Paper 4941[document on the Internet]. c2009 [cited 2014 July 25]. Available from: http://econ.worldbank.org
2. United Nations Development/ United Nations Capital Development Fund/ United Nations Environment (UNDP/UNCDF/UNEP). Local governance and climate change. A discussion note, December 2010. Bangkok: UNDP/ UNCDF/UNEP; 2010.
3. Omuko L. Climate change governance at subnational level: Key lessons for Kenya's county governments. Strathclyde Centre for Environmental Law and Governance (SCELG): Working paper 2/2015. Glasgow: SCELG; 2015.
4. Orlove B, Roncoli C, Kabugo M, Majugu A. Indigenous climate knowledge in southern Uganda: The multiple components of a dynamic regional system. Climate Change. 2010;100:243–265. http://dx.doi.org/10.1007/s10584-009-9586-2
5. Parry ML, Canziani OF, Palutikof JP, Van Der Linden PJ, Hanson CE, editors. Contribution of working group II to the fourth assessment report of the intergovernmental panel on climate change. Cambridge: Cambridge University Press; 2007. p. 976.
6. Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, et al, editors. Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2012. p. 582. http://dx.doi.org/10.1017/CBO9781139177245
7. Aggarwal A, Chhetri N, Cull T, Gustavo Feres J, Haggar J, Hutchinson G, et al, editors. Rural areas. Contribution of working group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2014. http://dx.doi.org/10.1017/CBO9781107415416.024
8. United Nations Environment Programme (UNEP). Indigenous knowledge in disaster management in Africa. Nairobi: UNEP; 2008.
9. United Nations International Strategy for Disaster Reduction (UNISDR). Indigenous knowledge for disaster risk reduction – good practices and lessons learned from experiences in the Asia-Pacific Region. Bangkok: UNISDR; 2008.
10. Gwimbi P. Linking rural community livelihoods to resilience building in flood risk reduction in Zimbabwe. Jamba. 2009;2(1):071–079. http://dx.doi.org/10.4102/jamba.v2i1.16
11. Carcellar N, Christopher J, Co R, Hipolito ZO. Addressing disaster risk reduction through community-rooted interventions in the Philippines: Experience of the Homeless People's Federation of the Philippines. Environ Urban. 2011;23(2):365–381. http://dx.doi.org/10.1177/0956247811415581
12. Satterthwaite D. Why is community action needed for disaster risk reduction and climate change adaptation? Environ Urban. 2011;23(2):339–349. http://dx.doi.org/10.1177/0956247811420009
13. Nyong A, Adesina F, Elasha BO. The value of indigenous knowledge in climate change mitigation and adaptation strategies in the African Sahel. Mitig Adapt Strat Glob Change. 2007;12:787–797. http://dx.doi.org/10.1007/s11027-007-9099-0
14. Gearheard S, Pocernich M, Stewart R, Sanguya J, Huntington HP. Linking Inuit knowledge and meteorological station observations to understand changing wind patterns at Clyde River, Nunavut. Climate Change. 2010;100(2):267–294. http://dx.doi.org/10.1007/s10584-009-9587-1
15. King DNT, Skipper A, Tawhai WB. Māori environmental knowledge of local weather and climate change in Aotearoa – New Zealand. Climate Change. 2008;90:385–409. http://dx.doi.org/10.1007/s10584-007-9372-y
16. Nakashima DJ, Galloway MK, Thulstrup HD, Ramos CA, Rubis JT. Weathering uncertainty: Traditional knowledge for climate change assessment and adaptation. Paris: United Nations Educational, Scientific and Cultural Organisation (UNESCO); 2012.
17. Salick J, Byg A. Indigenous peoples and climate change. Oxford: Tyndall Centre for Climate Change Research; 2007.
18. Berkes F. Indigenous ways of knowing and the study of environmental change. J Roy Soc New Zealand. 2009;39(4):151–156. http://dx.doi.org/10.1080/03014220909510568
19. Green D, Raygorodetsky G. Indigenous knowledge of a changing climate. Climate Change. 2010;100:239–242. http://dx.doi.org/10.1007/s10584-010-9804-y
20. World Bank. Indigenous knowledge: Local pathways to global development. Knowledge and Learning Group Africa Region [article on the Internet]. c2004 [cited 2014 July 20]. Available from: http://worldbank.org/afr/ik/default.htm
21. Welp M, De La Vega-Leinert A, Stoll-Kleemann S, Jaeger CC. Science-based stakeholder dialogues: Theories and tools. Glob Env Change. 2006;16:170–181. http://dx.doi.org/10.1016/j.gloenvcha.2005.12.002
22. International Union for Conservation of Nature (IUCN). Indigenous and traditional peoples and climate change. Issues Paper [document on the Internet]. c2008 [cited 2014 July 20]. Available from: http://cmsdata.iucn.org/downloads/indigenous_peoples_climate_change.pdf

23. United Nations Educational, Scientific and Cultural Organisation (UNESCO). Declaration on Science and the use of Scientific Knowledge and the Science Agenda – Framework of Action [document on the Internet]. c1999 [cited 2014 July 22]. Available from: http://unesco.at/wissenschaft/basisdokumente/about_wissenschaft.pdf

24. United Nations (UN). UN Conference on Environment and Development [document on the Internet]. c1992 [cited 2014 July 20]. Available from: http://un.org/geninfo/bp/enviro.html

25. United Nations (UN). Report of the World Commission on Environment and Development: Our Common Future [document on Internet]. c1987 [cited 2014 July 22]. Available from: http://un-documents.net/our-common-future.pdf

26. Biermann F, Boas I. Preparing for a warmer world: Towards a global governance system to protect climate refugees. Glob Environ Polit. 2010;10(1):60–88. http://dx.doi.org/10.1162/glep.2010.10.1.60

27. Ashton J, Wang X. Equity and climate in principle and practice. In: Bodansky D, Diringer E, Tudela F, Ashton J, Pershing J, Aldy JE, et al, editors. Beyond Kyoto: Advancing the international effort against climate change. Arlington: Pew Centre on Global Climate Change; 2003.

28. Andonova L, Betsill MM, Bulkeley H. Transnational climate change governance. Amsterdam Conference on the Human Dimensions of Global Environmental Change; 2007 May 24–26; Amsterdam, the Netherlands.

29. Backstrand K, Loubrand E. Climate governance beyond 2012: Competing discourses of green governmentality, ecological, modernisation and civic environmentalism. In: Pettenger ME, editor. The social construction of climate change: Power, knowledge, norms, discourses. Hampshire: Ashgate Publishing Ltd; 2007.

30. Pattberg P, Stripple J. Remapping global climate governance: Fragmentation beyond the public/private divide. Global governance working paper No. 32. Amsterdam: The Global Governance Project; 2007.

31. Van Asselt H. Dealing with the fragmentation of global climate governance. Legal and political approaches in interplay management. Global Governance Working Paper No. 30. Amsterdam: The Global Governance Project; 2007. http://dx.doi.org/10.2139/ssrn.1335082

32. Chirisa I, Chanza N. How will climate change transform African local governance? – assessing the role of civic engagement. J Public Admin Pol Res. 2009;1(2):035–046.

33. Bulkeley H, Moser SC. Responding to climate change: Governance and social action beyond Kyoto. Glob Environ Polit. 2011;7(2):1–10. http://dx.doi.org/10.1162/glep.2007.7.2.1

34. Saran S. Global governance and climate change. Glob Gov. 2009;15:457–460.

35. Martin S. Climate change, migration and governance. Glob Gov. 2010;16:397–414.

36. Oberthür S. Global climate governance after Cancun: Options for EU leadership. Int Spectator. 2011;46(1):5–13. http://dx.doi.org/10.1080/03932729.2011.567900

37. Pattberg P, Stripple J. Beyond the public and private divide: Remapping transnational climate governance in the 21st century. Int Environ Agreem-P. 2008;8:367–388. http://dx.doi.org/10.1007/s10784-008-9085-3

38. Okereke C, Bulkeley H. Conceptualizing climate change governance beyond the international regime. Tyndall Centre Working Paper, East Anglia: Tyndall Centre, University of East Anglia. Norwich: University of East Anglia; 2007.

39. Macey A. Climate change: Governance challenges for Copenhagen. Glob Gov. 2009;15:443–449.

40. Mawere M. Indigenous knowledge systems' (IKSs) potential for establishing a moral, virtuous society: Lessons from selected IKs in Zimbabwe and Mozambique. J Sust Dev Afr. 2010;12(7):209–221.

41. Mawere M. Culture, indigenous knowledge and development in Africa: Reviving interconnections for sustainable development. Bamenda: Langaa RPCIG Publishers; 2014.

42. Mawere M, Madziwa BF, Mabeza CM. Climate change and adaptation in third world Africa: A quest for increased food security in semi-arid Zimbabwe. Int J Human Soc Stud. 2013;1(2):14–22.

43. Shizha E. Reclaiming our indigenous voices: The problem with postcolonial sub-Saharan African school curriculum. J Indig Soc Dev. 2013;2(1):1–18.

44. Chanza N. Indigenous knowledge and climate change: Insights from Muzarabani, Zimbabwe [PhD Thesis]. Port Elizabeth: Nelson Mandela Metropolitan University; 2014.

45. Chanza N. Indigenous-based adaptation: An imperative for sustainable climate change strategies for Africa. In: Mawere M, Awuah-Nyamekye S, editors. Harnessing cultural capital for sustainability: A pan Africanist perspective. Bamenda: Langaa Publishing House; 2015. p.85–134.

46. Berkes F, Colding J, Folke C. Rediscovery of traditional ecological knowledge as adaptive management. Ecol Appl. 200;10(5):1251–1262.

47. Ajibade LT, Shokemi OO. Indigenous approaches to weather forecasting in Asa LGA, Kwara State, Nigeria. Indilinga: Afr J Indig Knowl Syst. 2003;2:37–44.

48. Brook RK, McLachlan SM. Trends and prospects for local knowledge in ecological and conservation research and monitoring. Biodivers Conserv. 2008;17:3501–3512. http://dx.doi.org/10.1007/s10531-008-9445-x

49. Mapara J. Indigenous knowledge systems in Zimbabwe: Juxtaposing postcolonial theory. J Pan Afr Stud. 2009;3(1):139–155.

50. Laborde S, Imberger J, Toussaint S. Contributions of local knowledge to the physical limnology of Lake Como, Italy. P Natl Acad Sci USA. 2012;109(17):6441–6445. http://dx.doi.org/10.1073/pnas.1113740109

51. Lefale PF. *Ua 'afa le Aso* Stormy weather today: Traditional ecological knowledge of weather and climate. The Samoa experience. Climate Change. 2010;100(2):317–335. http://dx.doi.org/10.1007/s10584-009-9722-z

52. Speranza CI, Kiteme B, Ambenje P, Wiesmann U, Mikal S. Indigenous knowledge related to climate variability and change: Insights from droughts in semi-arid areas of former Makueni District, Kenya. Climate Change. 2010;100:295–315. http://dx.doi.org/10.1007/s10584-009-9713-0

53. Makhado RA, Saidi AT, Mantlana BK, Mwayafu MD. Challenges of reducing emissions from deforestation and forest degradation (REDD+) on the African continent. S Afr J Sci. 2011;107(9/10), Art. #615, 3 pages. http://dx.doi.org/10.4102/sajs.v107i9/10.615

54. Richards M, Swan SR. Participatory subnational planning for REDD+ and other land use programmes: Methodology and step-by-step guidance. Ho Chi Minh: SNV Netherlands Development Organisation, REDD+ Programme; 2014.

55. United Nations Framework Convention on Climate Change (UNFCCC). Report of the conference of the parties serving as the meeting of the parties to the Kyoto Protocol on its first session; 2005 Nov 28 – Dec 10; Montreal, Canada [report of the Internet]. c2006 [cited 2014 May 07]. Available from: http://unfccc.int/resource/docs/2005/cmp1/eng/08a03.pdf

56. Mohamed-Katerere JC. Customary environmental management systems. In: Mohamed-Katerere JC, Chenje M, editors. Environmental law and policy in Zimbabwe. Harare: Southern African Research and Documentation Centre; 2002.

57. Turner NJ, Clifton H. ''It's so different today'': Climate change and indigenous lifeways in British Columbia, Canada. Glob Environ Chang. 2009;19:180–190. http://dx.doi.org/10.1016/j.gloenvcha.2009.01.005

58. Berkes F, Jolly D. Adapting to climate change: Social-ecological resilience in a Canadian western Arctic community. Conserv Ecol. 2001;5(2):18.

59. Adger WN, Huq S, Brown K, Conway D, Hulme M. Adaptation to climate change in the developing world. Prog Dev Studies. 2003;3:179–195. http://dx.doi.org/10.1191/1464993403ps060oa

60. Lavell A, Oppenheimer M, Diop C, Hess J, Lempert R, Li J, et al. Climate change: New dimensions in disaster risk, exposure, vulnerability, and resilience. In: Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, et al., editors. Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge: Cambridge University Press; 2012. p. 25–64. http://dx.doi.org/10.1017/CBO9781139177245.004

61. Ensor J. Governance for community-based adaptation. A practical action discussion paper. Rugby: Practical Action Publishing; 2009.

62. Sekine H, Fukuhara K, Uraguchi A, Tan CK, Nagai M, Okada Y. The effectiveness of community-based adaptation (CBA) to climate change – from the viewpoint of social capital and indigenous knowledge. Global Environment Information Centre (GEIC) working paper series 2009-001. Tokyo: GEIC; 2009.

63. Kirkland E. Indigenous knowledge and climate change adaptation in the Peruvian Andes [article on the Internet]. c2012 [cited 2014 July 28]. Available from: http://navsarjan.org/ids-document/indigenous-knowledge-and-climate-change-adaptation-in-the-peruvian-andes/

64. Munang R, Thiaw I, Alverson K, Mumba M, Liu J, Ravington M. Climate change and ecosystem-based adaptation: A new pragmatic approach to buffering climate change impacts. Curr Opin Environ Sustain. 2013;5(1):67–71. http://dx.doi.org/10.1016/j.cosust.2012.12.001

65. Colls A, Ash N, Ikkala N. Ecosystem-based adaptation: A natural response to climate change. Gland: International Union for Conservation of Nature (IUCN); 2009.

66. Boyd J. Ecosystem services and climate adaptation. Issue Brief 10-16. Washington, DC: Resources for the Future; 2010.

67. United Nations Framework Convention on Climate Change (UNFCCC). Ecosystem-based adaptation [document on the Internet]. c2012 [cited 2014 Aug 01]. Available from: http://unfccc.int/files/adaptation/application/pdf/nwp_cal_2012.pdf

68. Chanza N. Building social capital for sustainable rural development in Zimbabwe: Lessons from Korea's Saemaul Undong. Planning and Policy Report Vol. 2. Gyeonggi-do, Korea: Global Development Partnership Center; 2011; p. 84–101.

69. Woolcock M, Narayan D. Social capital: Implications for development theory, research and policy. World Bank Res Obs. 2000;15(2):225–249. http://dx.doi.org/10.1093/wbro/15.2.225

70. Climate Action Network International (CAN). An adaptation action framework of the Copenhagen agreement. Submission to the UNFCCC secretariat; 2009 Apr 24. Beirut: CAN, 2009.

Tuberculous meningitis in infants and children: Insights from nuclear magnetic resonance metabolomics

AUTHORS:
Shayne Mason[1]
Carolus J. Reinecke[1]
Regan Solomons[2]
A. Marceline van Furth[3]

AFFILIATIONS:
[1]Centre for Human Metabonomics, Faculty of Natural Sciences, North-West University, Potchefstroom Campus, Potchefstroom, South Africa
[2]Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa
[3]Department of Paediatric Infectious Diseases and Immunology, VU University Medical Center, Amsterdam, the Netherlands

CORRESPONDENCE TO:
Carolus Reinecke

EMAIL:
carools.reinecke@nwu.ac.za

POSTAL ADDRESS:
Centre for Human Metabonomics, Faculty of Natural Sciences, Private Bag X6001, Potchefstroom 2531, South Africa

DATES:
Received: 03 Mar. 2015
Revised: 03 Aug. 2015
Accepted: 31 Oct. 2015

KEYWORDS:
TBM; NMR spectroscopy; cerebrospinal fluid; CSF; hypothesis

HOW TO CITE:
Mason S, Reinecke CJ, Solomons R, Van Furth AM. Tuberculous meningitis in infants and children: Insights from nuclear magnetic resonance metabolomics. S Afr J Sci. 2016;112(3/4), Art. #2015-0086, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20150086

Tuberculous meningitis (TBM) is a prevalent form of central nervous system tuberculosis (CNS-TB) and the most severe common form of bacterial meningitis in infants and children below the age of 13 years, especially in the Western Cape Province of South Africa. Research to identify markers for timely and accurate diagnosis and treatment outcomes remains high on the agenda for TBM, in respect of which the field of metabolomics is as yet largely unexploited. However, the national Department of Science and Technology (DST) recently established several biotechnology platforms at South African institutions, including one for metabolomics hosted at North-West University. We introduce this national platform for nuclear magnetic resonance (NMR) metabolomics by providing an overview of work on TBM. We focus on selected collaborative multidisciplinary approaches to this disease and conclude with the outcomes of an untargeted NMR metabolomics study of cerebrospinal fluid from TBM patients. This study enabled the formulation of a conceptual shuttle representing the unique metabolic plasticity of CNS metabolism towards the energy requirements for the microglia-driven neuroinflammatory responses, of which TBM is one example. From insights generated by this explorative NMR metabolomics investigation, we propose directions for future in-depth research strategies to address this devastating disease. In our view, the timely initiative of the DST, the operational expertise in metabolomics now available and the potential for involving national and international networks in this field of research offers remarkable opportunities for the future of metabolomics in South Africa and for an ever greater understanding of disease mechanisms.

Introduction

Metabolomics has become an established scientific field. It provides analytical, chemical and physiological insights into naturally occurring, low-molecular-weight organic metabolites within organisms, cells, tissues and biofluids, and helps to promote understanding of metabolite interactions.[1] Explorative metabolomics investigations produce metabolite profiles and rely on the scientific method of induction to generate hypotheses in order to identify metabolites that may be indicators or biomarkers of disease or used as monitors of therapeutic responses.[2] Tuberculosis (TB), in its variety of manifestations – such as pulmonary and pleural TB, as well as TB within the central nervous system (CNS-TB), with tuberculous meningitis (TBM) as one of its most dangerous forms – is high on the list for such applications. Although still limited in number, metabolomic technologies have been applied successfully to diseases of the CNS. The composition of cerebrospinal fluid (CSF), which is partially derived from interstitial fluid in the CNS, is anticipated to reflect the normal and pathological biochemical processes of the brain. In the present review, we start by addressing the pathophysiology and severity of TBM and follow with a discussion on the application of metabolomics in enriching our understanding of this infectious disease. Research on TB, in its broadest sense, has a long-standing history in South Africa, emanating from several leading research units and centres, and directed and staffed by world-renowned scholars. Their contributions warrant a review alone and are not covered here; our main focus is on TBM, the latest pioneering investigations using nuclear magnetic resonance (NMR) metabolomics, and perspectives on methodological approaches that further exploratory insights and hypotheses formulated through metabolomics studies.

Tuberculosis is caused by the bacterium *Mycobacterium tuberculosis*. Although TB commonly presents as a pulmonary disease, systemic spread of the tubercule bacilli can lead to extrapulmonary forms of TB (EPTB), present in up to 20% of reported TB cases,[3] and particularly common in young children and immunocompromised individuals[4]. CNS-TB represents up to an estimated 10% of all EPTB (and 1% of total TB) cases.[3,5] According to the World Health Organization Global TB Report for 2014, there were 312 380 clinically defined new cases of TB in South Africa alone; of these, 37 709 (12%) were EPTB, which included an estimated 3771 cases of CNS-TB. Among the various manifestations of TB, TBM is not only the most prevalent form of CNS-TB but is also the most severe complication of TB and the most common form of bacterial meningitis in children below the age of 13 years[6], especially in the Western Cape Province of South Africa[7,8].

Recent reviews have covered various aspects of childhood TBM.[9-11] The general consensus emerging from these reports is that: (1) progress in understanding, prevention and treatment of TBM remains inadequate; (2) childhood deaths and disability as a result of TBM constitute a major societal burden; (3) although preliminary results obtained with the Bacillus Calmette–Guérin (BCG) vaccine seem to be promising for protection against TB, future trials of candidate vaccines are still needed; and (4) novel research to improve timely and accurate diagnosis and treatment outcomes remains high on the agenda. TBM is a disease that is largely unexploited by metabolomics technology, which predominantly makes use of hyphenated mass spectroscopy (MS) and NMR spectroscopy – the latter being the focus of this overview.

Pathogenesis and severity of TBM

Tuberculosis is spread via an aerosol route from persons who have active disease. Most people infected with *M. tuberculosis* have latent disease and are asymptomatic. However, a small proportion of individuals go on to develop active TB – with some developing systemic TB, such as TBM – and have significant morbidity and mortality, even though effective treatment is available.

In the late 19th century, it was thought that TBM resulted from haematogenous spread to the meninges, as a result of the frequent finding that TBM and miliary TB were occurring in the same patients.[12] In 1933 Rich and McCordock[13] published their report that, in the majority of post-mortems, a single caseous focus (Rich's focus) could be found, from which, when ruptured, bacilli could spread to the subarachnoid space. It was thought that this single lesion was almost always older than all the lesions found in concurrently occurring miliary TB, which set the paradigm for understanding the pathogenesis of TBM. However, this model did not fully explain the frequency of miliary TB and TBM occurring simultaneously[14], nor the mechanism whereby *M. tuberculosis* spreads from the lungs to the meninges and crosses the blood–brain barrier[15]. Animal models have been developed in order to understand better the pathogenesis of CNS-TB but findings are frustrated by the poor human clinical–pathological correlation.[15] Magnetic resonance imaging has detected numerous concurrent leptomeningeal granulomas in children with miliary TB, further challenging Rich's pathogenic model.[15,16] It is likely that early haematogenous spread to the brain occurs before a T-cell mediated immune response is activated. This mechanism could explain the vulnerability to TBM when T-cell mediated immunity is sub-optimal in persons infected with the human immunodeficiency virus (HIV)[9] and in persons on long-term glucocorticoid therapy[17]. The protective role of lymphocytes is essential, with the contribution of CD4+ and CD8+ T-cells, along with macrophages in isolating and engulfing *M. tuberculosis*; together, these eventually lead to granuloma formation. We, and others, have shown that many of the signs, symptoms and sequelae of TBM result from an immunologically directed inflammatory response to the infection.[18,19]

Greater understanding of the entry of *M. tuberculosis* into the CNS and the immunological mechanisms allowing survival of the bacilli is crucial for improving prevention and treatment. Fluctuation in cytokine levels affecting immunological function in patients with TBM can directly influence the duration of the disease and its severity.[18] In particular, the balance between pro- and anti-inflammatory cytokines may be crucial to TBM disease progression[19]; this relationship is reflected by the upregulation in CSF of pro- and anti-inflammatory cytokines in patients with TBM compared to patients with other forms of meningitis.[20,21] The early clinical presentation of TBM is often non-specific, with symptoms such as cough, loss of weight, fever, vomiting and malaise. As the disease progresses, meningism, focal neurological signs, and a depressed level of consciousness can occur. The timing of initiation of treatment is the most critical factor affecting morbidity, mortality, and healthcare costs, which emphasises the importance of early diagnosis of TBM.[22] Through a large retrospective study of 554 children with TBM, we established that 14% presented with visual impairment, 16% with hearing loss, 44% with motor deficits and 77% with intellectual impairment; only a small number (16%) did not have neurological manifestations of TBM.[23]

Accurate prediction of outcome in childhood TBM is difficult owing to the diversity of underlying pathological mechanisms and variation in host immunological response. Multidrug-resistant (MDR) TBM in children has a poor clinical outcome and is associated with increased mortality.[24] In-patient mortality rates do not differ between HIV-infected and non-HIV-infected children with TBM; however, mortality after hospital discharge is greater in HIV-infected children with TBM because of HIV-related illnesses.[25] Even though outcome prediction is problematic, the clinical stage of disease at the time of starting treatment is predictive of prognosis.[11] Children with Stage I TBM disease are likely to lead a normal life without neurological complications, whereas those with Stage III disease have a high risk of mortality.[11] Of all the TBM staging systems, the refined Medical Research Council scale has been shown to have the highest predictive value.[26,27] The criteria that define each stage are as follows: Stage I – Glasgow Coma Scale (GCS) of 15 and no focal neurology; Stage IIa – GCS of 15 plus focal neurology; Stage IIb – GCS of 11–14 with focal neurology; and Stage III – GCS <11.[26,27] The importance of early diagnosis corresponding to an early TBM stage is emphasised by the significantly increased risk, as the disease progresses, of severe motor deficit and degree of intellectual disability in children classified as Stage IIb compared to those classified as Stage IIa.[27] Stage III disease carries a 73% risk of the patient developing spastic quadriparesis and a mean developmental quotient of 44.[27] Besides the prediction of disease outcome according to TBM stage, convulsions, headache, motor deficit, brainstem dysfunction and cerebral infarctions on neuroimaging were independently associated with poor clinical consequences of TBM in multivariate analysis.[23]

CSF diagnostic indicators of TBM

Cerebrospinal fluid findings are integral to the diagnosis of TBM. Typical CSF findings consist of leukocytosis with lymphocyte predominance, elevated protein and abnormally decreased CSF glucose (hypoglycorrhachia).[28] Hypoglycorrhachia is indicated by either decreased CSF to plasma glucose ratio or reduced CSF glucose, and is unaffected by HIV co-infection.[29] A uniform research case definition for TBM identified CSF criteria to aid in the diagnosis of TBM, including a CSF to serum glucose ratio of less than 0.5 or an absolute CSF glucose concentration of less than 2.2 mmol/L and elevated CSF protein of greater than 1 g/L.

There is potential for the use of CSF lactate in children with TBM, as CSF lactate has been shown to differentiate between bacterial and aseptic meningitis.[30] CSF lactate levels, unaffected by serum lactate concentration, were significantly higher in adult TBM patients who subsequently died than in those who survived, reflecting the severity of cerebral hypoxia and therefore the overall prognosis.[31] The diagnostic utility of CSF indicators considered in TBM has, however, rarely been described. Studies evaluating CSF IGRA (interferon-gamma release assay) showed good sensitivity and specificity,[32] but the low CSF volumes in children are a limitation when sufficient cells are required to perform IGRA (typically 5–10 mL CSF is required).[11] A study of the host immune response to *M. tuberculosis* showed the potential value of CSF interleukin-13, vascular endothelial growth factor and cathelicidin LL-37 as indicators when differentiating TBM from other forms of meningitis.[18] Understanding the host immune response is key to a better understanding of the pathophysiology, the clinical presentation of TBM and the treatment of the disease. Novel indicators from CSF, including those from validated metabolomics studies, have the potential to be extremely useful as diagnostic tools in clinical practice.

NMR metabolomics methodology and applications

Metabolomics involves the simultaneous and comprehensive analysis of the measured responses of various biologically relevant small molecules, collectively referred to as the metabolome. The metabolome in turn is defined as the quantitative and qualitative collection of all low molecular weight molecules (metabolites, the end products of gene expression) of interest; they are found in concentrations varying in magnitude from picomoles to millimoles, they originate from endogenous or exogenous sources, and they exhibit a wide range of physico-chemical properties.[33] These small molecules display characteristics which allow metabolomics to avoid several challenges associated with genomics and proteomics investigations. The great chemical complexity of the metabolome, however, means that a single analytical technique is not sufficient to provide comprehensive characterisation of all metabolites.[33] Among the techniques available, NMR spectroscopy provides an attractive alternative to MS. Although it is less sensitive than its MS-based counterparts, NMR (1) allows for a robust and global look at the metabolome – it detects all classes of metabolites – making it non-biased; (2) is highly specific, allowing for detection of novel compounds; (3) running costs are markedly lower than that of other platforms, although initial set-up is expensive; (4) requires minimal sample preparation, and hence (5) does not chemically alter nor destroy the sample under investigation. NMR thus provides an excellent early or explorative view of the metabolome under investigation.[34]

The potential of NMR metabolomics in research on infectious human diseases has generated unique and novel insight into the underlying pathogenesis of these diseases and has unveiled new metabolic markers for disease diagnosis.[35] Infectious diseases relevant to Africa that have been studied using NMR metabolomics include human African trypanosomiasis[36], schistosomiasis[37], malaria[38], Leishmaniasis[39], pneumonia[40], and HIV[41], as well as TB and TBM (discussed below).

NMR metabolomics of TB and TBM

The potential of metabolomics has been well documented, specifically with regard to TB.[42] Its value is in the ability to (1) address simultaneously a wide array of metabolites from various different biochemical pathways; (2) detect and isolate patterns of disturbance for additional, targeted investigations; and (3) generate hypotheses to be tested. Godreuil et al.[43] believe that the new quantitative and bioinformatics approaches to the study of the interaction between *M. tuberculosis* and the infected host, and how this interaction influences the infection process, are of particular importance, as it is fully accepted in the scientific description of infectious diseases that the outcome of transmission, infection and disease is dependent on the intrinsic characteristics of both the microbes and the host.

It is expected that at different stages of the *M. tuberculosis* life cycle – for example, dormancy, latent infection and active disease – there will be a different array of host- and *M. tuberculosis*-derived metabolites.[42] These metabolic profiles can best be measured globally via NMR metabolomics (Table 1), which has been used successfully to provide insight into the metabolic changes in host response. In a pioneering NMR-based metabolomics pilot study of the CSF of patients with bacterial meningitis, Coen et al.[47] detected: (1) elevated lactate and severe glucose depletion; (2) impairment of the citric acid cycle caused by reduced production of acetyl CoA, resulting in accumulation of pyruvate and generation of amino acids via transamination; (3) elevated CSF concentrations of pyruvate and amino acids – particularly alanine, isoleucine and leucine; and (4) the presence of ketone bodies – 3-hydroxybutyrate, acetoacetate and acetone – indicating compensatory response to glucose depletion and reduced ATP levels. Notably, the role of amino acid perturbation in meningitis is further supported by other studies.[48] The metabolite profile described by Coen et al.[47] for bacterial meningitis corresponds to recent MS-based results from our research centre on CSF from a TBM patient[49] and to sputum from patients with pulmonary TB[50]. The pilot study of Coen et al.[47] thus highlights the potential of NMR metabolomics in providing a global and unbiased view of metabolic or neurological perturbations in infectious diseases, particularly meningitis.

A study by Zhou et al.[45] unambiguously identified 35 metabolites in sera of TB patients, of which 17 were altered. The majority of altered endogenous metabolites in the serum of TB patients were energy related – citric acid cycle intermediates, products of glycolysis, amino acids and indicators of enhanced lipid degradation – which is consistent with other studies. The metabolic processes found to be most significantly altered in TB patients were protein biosynthesis, followed by alanine metabolism, phenylalanine and tyrosine metabolism, and ammonia recycling. Zhou et al.[45] go on to state that increased glycolytic products could be an index of tissue hypoxia and extent of necrosis as the infection progresses, that enhanced lipid degradation can be correlated with caseation of host TB granulomas, and that increased formate reflects an increased requirement for nucleotide biosynthesis, all of which indicates active host inflammatory cell division. A follow-up study by Zhou et al.[46] highlighted the specificity of the metabolite profile of TB patients, compared to similar diseases such as representative metabolism-related diseases (diabetes mellitus), wasting diseases (malignancy), and lung inflammatory diseases (CAP). Consistent with their previous study, they found in the plasma: increased levels of lactate, pyruvate, lipids and ketone bodies, and decreased levels of glucose, glutamate, glutamine, glycerophosphocholine, very low-density lipoproteins and branched-chain amino acids. While Zhou et al.[46] could distinguish TB patients statistically, the greatest statistical overlap occurred with the plasma profiles of patients with malignancy, indicating that, to some extent, the metabolism of TB mimics that of tumour cells. NMR metabolomics has also been used to elucidate the mechanism of action of specific medication, such as the second-line drug for TB, D-cycloserine, which is primarily used to treat MDR-TB.[51]

Table 1: Insights offered by nuclear magnetic resonance (NMR)-based metabolomics studies specific to tuberculosis (TB) and meningitis

Pathogen	Disease	Experimental subjects	Insights	Reference
Mycobacterium tuberculosis	Tuberculosis	Adults	Patients with TB were distinguished from healthy controls. Metabolic profile of sera of patients with TB indicated significant dysregulation of metabolic pathways, validating metabolic profiling of *M. tuberculosis*-infected murine models (Shin et al.[44]). Potential to develop novel clinical tools for TB diagnosis or therapeutic monitoring and improve understanding of disease mechanisms.	Zhou et al.[45]
M. tuberculosis	Tuberculosis	Adults	Plasma metabolite profile of patients with TB exhibited specificity when compared to representative metabolism-related diseases (diabetes mellitus), wasting diseases (malignancy), and lung inflammatory diseases (CAP). The closest comparison to the metabolism profile of TB was that of malignant tumours.	Zhou et al.[46]
Various	Bacterial meningitis	Children and adults	NMR-based analysis of cerebrospinal fluid (CSF) is feasible and a potentially more powerful diagnostic tool than conventional rapid laboratory indicators for distinguishing bacterial from viral meningitis (as well as distinguishing between species of bacteria), with important implications for early management, reduced empirical use of antibiotics and treatment duration.	Coen et al.[47]
M. tuberculosis	Tuberculous meningitis (TBM)	Infants and children	Differentiation between TBM and non-TBM cases based on 16 NMR-derived CSF metabolites – perturbed glucose, highly elevated lactate and several energy-related metabolites, signalling components and amino acid shuttles. Postulated the following hypothesis: 'The host's response to neural infection results in an "astrocyte–microglia lactate shuttle" that operates in neuroinflammatory diseases, such as TBM.'	Mason et al.[52]
Various	Meningitis	Children	Integration of NMR spectral information with routine clinical features that are incorporated into an in-house software system which allows for the rapid differential diagnosis of meningitis (bacterial, TB and viral).	Subramanian et al.[53]

We have recently used a gas chromatography–mass spectrometry (GC–MS) approach to investigate the analytical repeatability of a CSF sample collected from a TBM patient for metabolomics analysis[49] in order to develop a new method for evaluating repeatability in generating metabolomics data. This study clearly indicated the severe disease state of the TBM patient from whom the sample was obtained. We subsequently embarked on a comprehensive untargeted proton magnetic resonance (^{1}H NMR)-based metabolomics analysis on lumbar CSF samples from three experimental groups: (1) South African infants and children with confirmed TBM; (2) treated, non-meningitis South African infants and children as controls; and (3) age-matched untreated controls from the Netherlands.[52] Our primary focus was on identification of metabolites that distinguish between TBM and non-TBM cases; a total of 16 NMR-derived CSF metabolites was identified. These metabolites did not include cyclopropane, a putative marker for TBM defined in 2005.[53] The defining two metabolites were the combination of perturbed glucose and highly elevated lactate, common to some other neurological disorders. Glucose, the primary source for energy production, is utilised in a rapid fashion, resulting in a significant increase in CSF lactate, both of which reflect a metabolic burst. The remaining 14 metabolites of the host's response to TBM were likewise mainly energy-associated indicators. Citrate and pyruvate, along with acetate in the form of acetyl-CoA, are key metabolites involved in the citric acid cycle as part of energy production. Alanine and the branched-chain amino acids – valine, leucine and isoleucine – have vital roles as shuttling systems transporting metabolites across cell membranes; they are also anaplerotic, replenishing citric acid cycle intermediates. Myo-inositol acts as an important signalling component necessary for immune responses such as microglia activation. The remaining other metabolites also have important roles; one such is dimethyl sulfone, which is an indicator of response to oxidative stress; another is choline, an important precursor to acetylcholine, which in excess is responsible for seizures.

The most conspicuous outcome of the NMR profiling was the high concentration of lactate in the TBM patients (7.36 mmol/L) relative to the two control groups (1.39 mmol/L and 1.70 mmol/L for the Dutch and South African controls, respectively). This observation has some clinical significance. It has been established[54] that CSF lactate is applicable as a marker that can differentiate bacterial meningitis (>6 mmol/L) from other forms of meningitis, such as aseptic meningitis (<2 mmol/L). Thus, CSF concentrations of lactate, as well as immune biomarkers, are typically high before treatment and then decrease rapidly with therapy. In this regard a study[55] on adults suffering from meningitis indicated that death was associated with high initial CSF concentrations of lactate and low numbers of white blood cells, in particular neutrophils. This outcome underscores the authors' proposal that CSF lactate is a good single indicator of the TBM disease state, making it better than conventional markers.

Furthermore, our NMR-profiling results[52] provided a basis for hypothesis formulation of the pathogenic characteristics of TBM. We speculated that the inflammatory responses and metabolic imbalances created in the CNS following the initial phases of infection by a pathogen, such as *M. tuberculosis*, should advantage the microglia (the native macrophage in the brain) by fulfilling their energy-intensive immune-protective function. This speculation was made against the multi-paradigm background of the current literature and based on: (1) the shift in energy requirements as a result of the neuroinflammatory responses in TBM; (2) the notion that microglia play an important role in the neuropathogenesis of TBM; (3) recognition of the unique metabolic plasticity of cell–cell communication and regulation in CNS metabolism; and (4) the clinical profile on dysfunctional neuron activity seen in TBM. From observations and through inductive reasoning[56] on the characteristics of the cell–cell interactions and factor isoforms, we advanced the following hypothesis: 'The host's response to neural infection results in an "astrocyte–microglia lactate shuttle" (AMLS) that operates in neuroinflammatory diseases, such as TBM', represented as a conceptual model detailed in the supplementary material of Mason et al.[52] A simplified representation of the conceptual model is presented in Figure 1.

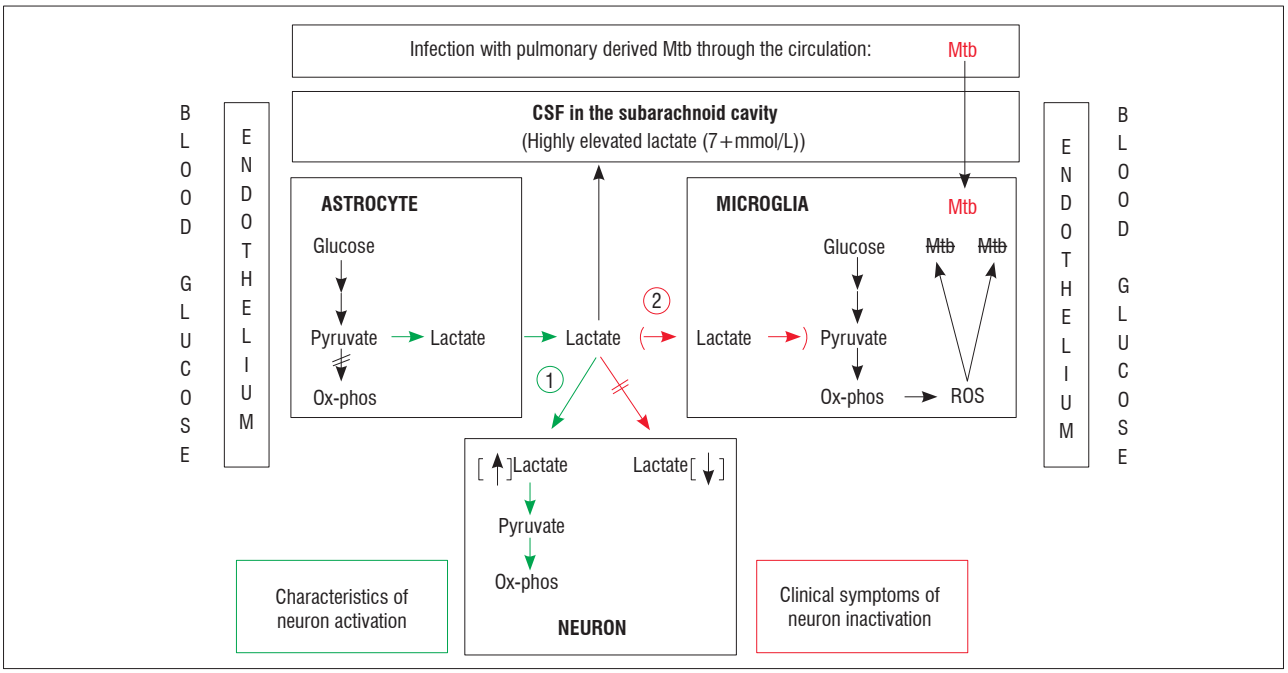


Figure 1: Representation of metabolic pathways of two lactate shuttles within the central nervous system. (1) The astrocyte–neuron lactate shuttle (ANLS; green pathway)[58] is operative under normal physiological conditions with astrocytes responding to glutamatergic activation by increasing their rate of glucose utilisation and release of lactate in the extracellular space, making the lactate available for neurons to sustain their energy demands. (2) The hypothetical astrocyte–microglia lactate shuttle (AMLS; red pathway)[52] proposed for tuberculous meningitis (TBM), with astrocytes responding to signalling from *Mycobacterium tuberculosis* (Mtb)-infected microglia by increasing glucose mobilisation, leading to increased extracellular lactate, reflected in increased cerebrospinal fluid (CSF) lactate levels, available as an additional energy source for reactive oxygen species (ROS) production aimed at destroying the invading Mtb.

Briefly, it is postulated that, in TBM, lactate produced through glycolysis in astrocytes participates in the activated immune response and is directed from the neurons preferentially into microglia, where it enters the mitochondrial citric acid cycle, contributing to oxidative phosphorylation and hence producing high levels of adenosine triphosphate (ATP) and formation of reactive oxygen species (ROS) required for *M. tuberculosis* degradation. ROS, and a multitude of factors produced by the microglia to modulate the functions of surrounding immune cells, are toxic to neurons. Moreover, the unregulated activation of microglia in response to stimulants such as *M. tuberculosis* propagate neural injury[57] and eventual apoptotic cell death for the over-activated microglia. By contrast, several investigations indicate that astrocytes could release large amounts of lactate under conditions of high energy demand by the neurons, leading to our hypothesis of an activity-dependent astrocyte–neuron lactate shuttle for the supply of energy substrates to neurons.[58]

From the preceding overview we highlight the following insights that have transpired from the NMR metabolomics of TB and TBM.

Advantages of NMR technology for clinical applications

(1) Non-invasive sample collection methods are often essential in clinical medicine, for which body fluids such as urine, cord blood and serum/plasma are particularly suited, with NMR spectroscopy providing a cornerstone for their metabolomics investigations. In addition, NMR spectroscopy can work with small sample volumes.

(2) NMR analyses require a minimum of sample pre-preparation and thereby prevent the occurrence of confounding analytical artefacts.

(3) Metabolomics produces large and complex data sets that cannot be interpreted through classic reductionist methods, but require an inductive, open-minded approach, aimed at hypothesis generation, which is philosophically a useful methodological approach in uncharted fields of research.

(4) Clinically, perturbations in metabolism are often difficult to recognise as the phenotypes may show considerable variation; however, NMR technology has provided a proven track record through the identification of inborn errors of metabolism, often already manifesting in neonates.

(5) The unique advantage of metabolomics technology is that it provides a dynamic view of host functional responses in health and disease, and offers early and rapid identification of the status and progress of a disease.

Cost-effectiveness of NMR technology

High-resolution NMR equipment, such as the instrument installed in early 2015 in a collaboration between the Technological Innovation Agency (TIA) and North-West University (NWU), is extremely expensive and requires highly skilled technical and scientific expertise for its operation. However, once an NMR facility has been established, the analytical clinical analysis is relatively inexpensive, costing only a few rands per sample to run. Moreover, the international trend in the market for miniaturisation of NMR instrumentation is well established, and is opening up new areas of study in academic and industrial settings with clear potential and appeal for future clinical practice.[59]

Systems insights

Fields that are data rich but hypothesis poor – prime examples are systems biology and many areas of medicine and clinical practice – are best approached through data-driven, essentially inductive methods. Biological systems are inherently complex; they require an inductive approach to data generation, data analysis and modelling if we are to continue to make strides in our understanding of these phenomena in health and disease. The advent of multidisciplinary interaction between biologists/clinicians, analytical chemists and informaticists/biostatisticians in developing complex models has ushered in a new era in the mathematics of science, for the first time enabling understanding and prediction of large-scale biological systems; the 'omics' technologies are key in this process by being inductive and exploratory in their very essence. Intrinsic to this process remains the iterative relationship between the realm of ideas/hypotheses/thoughts (= induction) and that of observations/data/information (= deduction), linked in a repetitive cyclic way in which one arc is not simply the reverse of the other, but has methodological characteristics in its own right. We conclude our overview by focusing on this cyclic relationship.

Directives for hypothesis verification

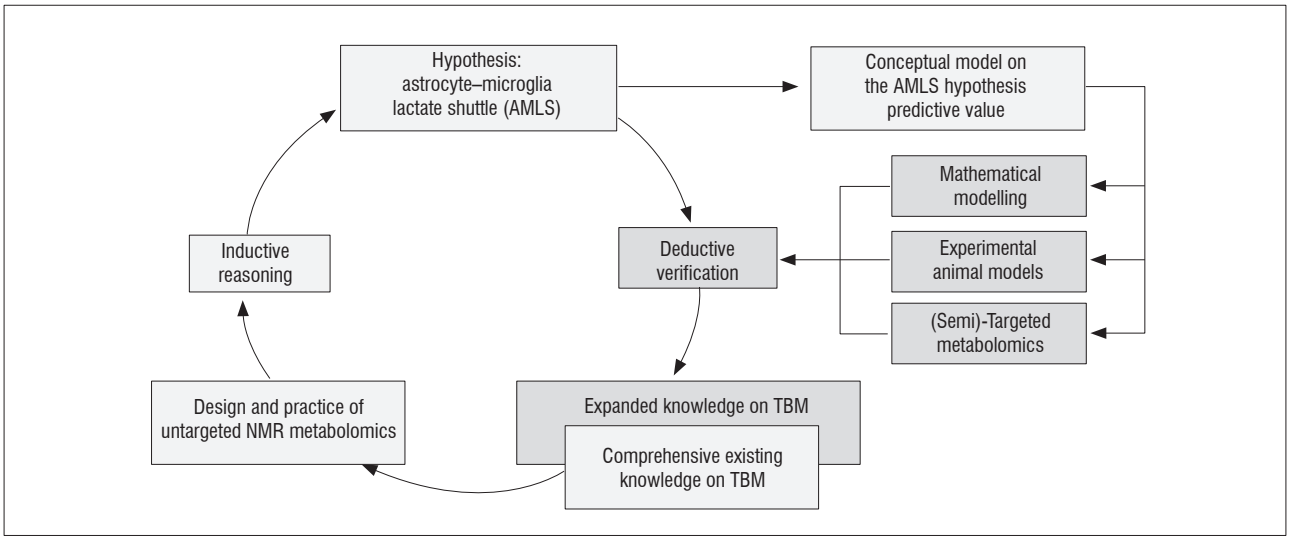
Hypotheses are generalised statements derived from empirical evidence and prior knowledge of existing data, and they need to be verified or refuted by deductive methods (Figure 2). This activity can be represented as a circular process, with inductive reasoning forming one half of the iterative cycle of knowledge, and the second half (deduction) devoted to the aims of testing hypotheses and furthering existing knowledge. This representation highlights the essential nature of exploratory metabolomics investigations, and features three approaches that could be applied for hypothesis testing: (1) computational models, (2) animal models and (3) semi-targeted metabolomics.

Computational modelling

The extreme complexity of intracellular regulatory networks, involving genetic circuits, metabolic regulation and signal transduction, hampers the ability to analyse these functions quantitatively to achieve a holistic understanding of their role in health and disease. Complexity theory has thus become established over the last two decades as a novel discipline directed to such problems in the humanities and natural sciences, and is at present pursued in South Africa at the Centre for Studies in Complexity established in 2009 at Stellenbosch University. Modelling, in its various forms, demands key competencies in studying complexity. Models are used to describe complex systems through computer simulation based on quantitative information on many and key variables that characterise what is being investigated. An example is the computational model we recently developed – using in silico data that capture the unique immunological environment of the brain – which allows us to study the key mechanisms driving TB-linked granuloma formation in time.[60] The model was based on a known pulmonary agent-based description of TB, representing the molecular signalling networks that affect granuloma formation during infection with *M. tuberculosis*. It focused on the formation of granulomas whose structure and function may reflect the success or failure of the host to contain infection. We envisage that computational model development, based on NMR data, may be a fruitful approach to use for understanding the dynamics involved in metabolic flux associated with the AMLS hypothesis, thereby permitting greater comprehension of the perturbations following *M. tuberculosis* infection of the microglia in the meninges.

Experimentation using animal models

Animal models have contributed significantly to knowledge of TB. However, there are certain major differences between humans and other animals in their responses to the disease. It is known that mice, for example, do not form the granulomas seen in humans and most other vertebrates with TB. Such differences caution our interpretations of data from animal models as they apply to human diseases. Nonetheless, animal models are indispensable to TB research and are well established in South African TB research centres.[61,62] Until recently, for example, no murine model was available for experimental studies on TBM. We successfully developed a reproducible in vivo murine model to study the inflammatory response in TBM,[63] and obtained useful insights despite limitations in translating the experimental observations to the human manifestation of the disease. Although the route of infection mimics the way TBM is acquired naturally, the cytokine profile in the mice does not show the clinical characteristics of the human disease. Nonetheless, the model has potential for further experimentation on *M. tuberculosis* infection because we observed that bacterial growth of TB in the CNS leads to a typical chronic inflammatory response. Such results indicate that experimental animal studies hold potential for investigating the biological changes in a *M. tuberculosis*-infected model at the biochemical level – an approach that could be expanded through appropriate knock-out models, as has been shown for cryptococcal meningitis.[64]



TBM, tuberculous meningitis; NMR, nuclear magnetic resonance

Figure 2: Iterative cycle of knowledge, using the astrocyte–microglia lactate shuttle (AMLS) hypothesis as an example. The existing knowledge forms the basis of inductive reasoning that leads to the formulation of hypotheses, which in turn is followed by the use of deduction to verify these hypotheses and to further existing knowledge.

Semi-targeted metabolomics

The metabolites present in the CSF represent the actual metabolism of the CNS and the balance between blood and CSF, rendering CSF analysis indispensable in the assessment of neurological disorders. Sampling human CSF is, however, an invasive procedure and ethical and safety constraints limit the availability of such samples for TBM, especially in work with infants and children. Thus, accessing these samples is much more difficult than collecting blood or urine. Metabolomics offers promising options. The use of MS-based metabolomics of plasma from patients suffering from propionic acidemia and methylmalonic acidemia,[65] as well as our GC-MS metabolomics analysis of urine samples from patients with respiratory chain deficiencies[66] and isovaleric acidemia[67], have provided proof of concept that semi-targeted metabolomics can expand the range of metabolites associated with human disease. These studies also indicate the potential of metabolomics for non-invasive diagnosis and clinical evaluation of patients with TBM.

Perspective

Metabolomics studies of complex diseases are still in their infancy, although the volume of NMR- and MS-based metabolomics publications in particular is increasing exponentially.[2] As it stands, NMR metabolomics is not a standalone method of diagnosis nor, by extension, of disease management of TB or of TBM, but it undeniably offers insights for potentially improving existing knowledge of infectious diseases in ways that may complement and benefit current diagnostic and treatment methods. Recent results, such as those by Zhou et al.[46], strongly indicate that NMR metabolomics can contribute to improved understanding of disease mechanisms. However, a major gap in work on infectious diseases is still the paucity of subsequent validation of clinical studies, which is necessary to reap the full rewards of metabolomics technology. Ultimately, the optimal outcome, particularly in resource-limited settings, is the validation and declaration of simple and clear biomarkers for disease diagnosis – which requires appropriate and thorough follow-ups on metabolomics studies.

Fortunately, the great potential of biotechnology was recognised by the Department of Science and Technology, which enabled the creation of several technological platforms, including the Metabolomics Platform supported by the TIA and hosted by NWU on its Potchefstroom campus. This platform consists of a wide array of gas and liquid chromatography based MS instruments, as well as a 500-MHz NMR spectrometer dedicated to investigations of biofluids and other biological specimens. Furthermore, the contractual agreement between the TIA and NWU stipulates that: '...the Institution [NWU] agrees to have an open policy that encourages use of its facilities by researchers other than those based at the Institution…'. This policy of open access to the Metabolomics Platform aligns with the TIA's objective of stimulating and exploiting technological innovation in order to improve economic growth and the quality of life of all South Africans – a view that we strongly share, with the hope that this review will further promote these far-reaching ambitions.

Acknowledgements

Research funding for this project was provided by the Technological Innovation Agency of the Department of Science and Technology of South Africa. S.M. and R.S. are recipients of a Desmond Tutu–NRF–VU doctoral fellowship for a joint PhD study between the Vrije Universiteit in Amsterdam (the Netherlands), and North-West University and Stellenbosch University in South Africa.

Authors' contributions

S.M. and R.S. wrote the manuscript. C.J.R. and A.M.v.F. are the South African and Dutch project leaders, respectively, who directed the study and contributed to the writing of the manuscript.

References

1. Lindon JC, Nicholson JK, Holmes E. The handbook of metabonomics and metabolomics. London: Elsevier Science; 2011.
2. Goodacre R. An overflow of… what else but metabolism! Metabolomics. 2010;6:1–2. http://dx.doi.org/10.1007/s11306-010-0201-3
3. Rieder HL, Kelly GD, Bloch AB, Cauthen GM, Snider DE. Tuberculosis diagnosed at death in the United States. Chest. 1991;100:678–681. http://dx.doi.org/10.1378/chest.100.3.678
4. Perez-Velez CM, Marais BJ. Tuberculosis in children. New Engl J Med. 2012;367(4):348–361. http://dx.doi.org/10.1056/NEJMra1008049
5. Bhigjee AI, Padayachee R, Paruk H, Hallwirth-Pillay KD, Marais S, Connoly C. Diagnosis of tuberculous meningitis: Clinical and laboratory parameters. Int J Infect Dis. 2007;11:248–254. http://dx.doi.org/10.1016/j.ijid.2006.07.007
6. Wolzak NK, Cooke ML, Orth H, Van Toorn R. The changing profile of pediatric meningitis at a referral centre in Cape Town, South Africa. J Trop Pediatr. 2012;58(6):491–495. http://dx.doi.org/10.1093/tropej/fms031

7. Donald PR, Cotton MF, Hendricks MK, Schaaf HS, De Villiers JN, Willemse TE. Pediatric meningitis in the Western Cape Province of South Africa. J Trop Pediatr. 1996;42:256–261. http://dx.doi.org/10.1093/tropej/42.5.256

8. Van Rie A, Beyers N, Gie RP, Kunneke M, Zietsman L, Donald PR. Childhood tuberculosis in an urban population in South Africa: Burden and risk factor. Arch Dis Child. 1999;80:433–437. http://dx.doi.org/10.1136/adc.80.5.433

9. Thwaites GE, Van Toorn R, Schoeman J. Tuberculous meningitis: More questions, still too few answers. Lancet Neurol. 2013;12:999–1010. http://dx.doi.org/10.1016/S1474-4422(13)70168-6

10. Chiang SS, Khan FA, Milstein MB, Tolman AW, Benedetti A, Starke JR, et al. Treatment outcomes of childhood tuberculous meningitis: A systematic review and meta-analysis. Lancet Infect Dis. 2014;14:947–957. http://dx.doi.org/10.1016/S1473-3099(14)70852-7

11. Van Toorn R, Solomons R. Update on the diagnosis and management of tuberculous meningitis in children. Semin Pediatr Neurol. 2014;21:12–18. http://dx.doi.org/10.1016/j.spen.2014.01.006

12. Hektoen L. The vascular changes of tuberculous meningitis, especially tuberculous endarteritis. J Exp Med. 1896;1:112–163. http://dx.doi.org/10.1084/jem.1.1.112

13. Rich AR, McCordock HA. The pathogenesis of tuberculous meningitis. Bull Johns Hopkins Hosp. 1933;52:2–37.

14. Donald PR, Schaaf HS, Schoeman JF. Tuberculous meningitis and miliary tuberculosis: The Rich focus revisited. J Infect. 2005;50(3):193–195. http://dx.doi.org/10.1016/j.jinf.2004.02.010

15. Thwaites GE, Schoeman JF. Update on tuberculosis of the central nervous system: Pathogenesis, diagnosis, and treatment. Clin Chest Med. 2009;30(4):745–754. http://dx.doi.org/10.1016/j.ccm.2009.08.018

16. Janse van Rensburg P, Andronikou S, Van Toorn R, Pienaar M. Magnetic resonance imaging of miliary tuberculosis of the central nervous system in children with tuberculous meningitis. Pediatr Radiol. 2008;38(12):1306–1313. http://dx.doi.org/10.1007/s00247-008-1028-1

17. Horsburgh CR Jr, Rubin EJ. Clinical practice. Latent tuberculosis infection in the United States. New Engl J Med. 2011;364(15):1441–1448. http://dx.doi.org/10.1056/NEJMcp1005750

18. Visser DH, Solomons RS, Ronacher K, Van Well GT, Heymans MW, Walzl G, et al. Host immune response to tuberculous meningitis. Clin Infect Dis. 2014;60(2):177–187. http://dx.doi.org/10.1093/cid/ciu781

19. Kashyap RS, Deshpande PS, Ramteke SR, Panchbhai MS, Purohit HJ, Taori GM, et al. Changes in cerebrospinal fluid cytokine expression in tuberculous meningitis patients with treatment. Neuroimmunomodulation. 2010;17:333–339. http://dx.doi.org/10.1159/000292023

20. Patel VB, Singh R, Connolly C, Kasprowicz V, Ndung'u T, Dheda K. Comparative utility of cytokine levels and quantitative RD-1-specific T cell responses for rapid immunodiagnosis of tuberculous meningitis. J Clin Microbiol. 2011;49(11):3971–3976. http://dx.doi.org/10.1128/JCM.01128-11

21. Donald PR, Schoeman JF, Beyers N, Nel ED, Carlini SM, Olsen KD, et al. Concentrations of interferon γ, tumor necrosis factor α, and interleukin-1β in the cerebrospinal fluid of children treated for tuberculous meningitis. Clin Infect Dis. 1995;21(4):924–929. http://dx.doi.org/10.1093/clinids/21.4.924

22. Schoeman JF, Wait J, Burger M, Van Zyl F, Fertig G, Janse van Rensburg A, et al. Long-term follow-up of childhood tuberculous meningitis. Dev Med Child Neurol. 2002;44(8):522–526. http://dx.doi.org/10.1111/j.1469-8749.2002.tb00323.x

23. Van Well GT, Paes BF, Terwee CB, Springer P, Roord JJ, Donald PR, et al. Twenty years of pediatric tuberculous meningitis: A retrospective cohort study in the Western Cape of South Africa. Pediatrics. 2009;123(1):e1–e8. http://dx.doi.org/10.1542/peds.2008-1353

24. Seddon JA, Visser DH, Bartens M, Jordaan AM, Victor TC, Van Furth AM, et al. Impact of drug resistance on clinical outcome in children with tuberculous meningitis. Pediatr Infect Dis J. 2012;31(7):711–716. http://dx.doi.org/10.1097/INF.0b013e318253acf8

25. Van Toorn R, Schaaf HS, Laubscher JA, Van Elsland SL, Donald PR, Schoeman JF. Short intensified treatment in children with drug-susceptible tuberculous meningitis. Pediatr Infect Dis J. 2014;33(3):248–252. http://dx.doi.org/10.1097/INF.0000000000000065

26. Van Toorn R, Springer P, Laubscher JA, Schoeman JF. Value of different staging systems for predicting neurological outcome in childhood tuberculous meningitis. Int J Tuberc Lung Dis. 2012;16(5):628–632.

27. British Medical Research Council. Streptomycin treatment of tuberculous meningitis. Br Med J. 1948;1(6503):582–596.

28. Thwaites G, Fisher M, Hemingway C, Scott G, Solomon T, Innes J. British Infection Society guidelines for the diagnosis of tuberculosis of the central nervous system in adults and children. J Infect. 2009;59:167–187. http://dx.doi.org/10.1016/j.jinf.2009.06.011

29. Marais S, Thwaites G, Schoeman JF, Török ME, Misra UK, Prasad K, et al. Tuberculous meningitis: A uniform case definition for use in clinical research. Lancet Infect Dis. 2010;10(11):803–812. http://dx.doi.org/10.1016/S1473-3099(10)70138-9

30. Huy NT, Thao NT, Diep DT, Kikuchi M, Zamora J, Hirayama K. Cerebrospinal fluid lactate concentration to distinguish bacterial from aseptic meningitis: A systemic review and meta-analysis. Crit Care. 2010;14(6), R240, 15 pages. http://dx.doi.org/10.1186/cc9395

31. Thwaites GE, Simmons CP, Than Ha Quyen N, Thi Hong Chau N, Phuong Mai P, Thi Dung N, et al. Pathophysiology and prognosis in Vietnamese adults with tuberculous meningitis. J Infect Dis. 2003;188:1105–1115. http://dx.doi.org/10.1086/378642

32. Thwaites GE. Advances in the diagnosis and treatment of tuberculous meningitis. Curr Opin Neurol. 2013;26(3):295–300. http://dx.doi.org/10.1097/WCO.0b013e3283602814

33. Dunn WB, Ellis DI. Metabolomics: Current analytical platforms and methodologies. Trends Anal Chem. 2005;24(4):285–294. http://dx.doi.org/10.1016/j.trac.2004.11.021

34. Kell DB, Oliver SG. Here is the evidence, now where is the hypothesis? The complementary roles of inductive hypothesis-driven science in the post-genomics era. BioEssays. 2003;66:99–105.

35. Duarte IF, Diaz SO, Gil AM. NMR metabolomics of human blood and urine in disease research. J Pharm Biomed Anal. 2014;93:17–26. http://dx.doi.org/10.1016/j.jpba.2013.09.025

36. Wang Y, Utzinger J, Saric J, Li JV, Burckhardt J, Dirnhofer S, et al. Global metabolic responses of mice to *Trypanosoma brucei brucei* infection. Proc Natl Acad Sci USA. 2008;105(16):6127–6132. http://dx.doi.org/10.1073/pnas.0801777105

37. Wang Y, Holmes E, Nicholson JK, Cloarec O, Chollet J, Tanner M, et al. Metabonomic investigations in mice infected with *Schistosoma mansoni*: An approach for biomarker identification. Proc Natl Acad Sci USA. 2004;101(34):12676–12681. http://dx.doi.org/10.1073/pnas.0404878101

38. Li JV, Wang Y, Saric J, Nicholson JK, Dirnhofer S, Singer BH, et al. Global metabolic responses of NMRI mice to an experimental *Plasmodium berghei* infection. J Proteome Res. 2008;7(9):3948–3956. http://dx.doi.org/10.1021/pr800209d

39. Gupta N, Goyal N, Singha UK, Bhakuni V, Roy R, Rastogi AK. Characterization of intracellular metabolites of axenic amastigotes of *Leishmania donovani* by ^{1}H NMR spectroscopy. Acta Trop. 1999;73(2):121–133. http://dx.doi.org/10.1016/S0001-706X(99)00020-0

40. Slupsky CM, Rankin KN, Fu H, Chang D, Rowe BH, Charles PG, et al. Pneumococcal pneumonia: Potential for diagnosis through a urinary metabolic profile. J Proteome Res. 2009;8(12):5550–5558. http://dx.doi.org/10.1021/pr9006427

41. Hewer R, Vorster J, Steffens FE, Meyer D. Applying biofluid ^{1}H NMR-based metabonomic techniques to distinguish between HIV-1 positive/AIDS patients on antiretroviral treatment and HIV-1 negative individuals. J Pharm Biomed Anal. 2006;41(4):1442–1446. http://dx.doi.org/10.1016/j.jpba.2006.03.006

42. Parida SK, Kaufmann SHE. The quest for biomarkers in tuberculosis. Drug Discov Today. 2010;15(3/4):148–157. http://dx.doi.org/10.1016/j.drudis.2009.10.005

43. Godreuil S, Tazi L, Bañuls AL. Pulmonary tuberculosis and mycobacterium tuberculosis: modern molecular epidemiology and perspectives. In: Tibayrenc M, editor. Encyclopedia of infectious diseases: Modern methodologies. Hoboken, NJ: John Wiley & Sons; 2007. http://dx.doi.org/10.1002/9780470114209.ch1

44. Shin JH, Yang JY, Jeon BY, Yoon YJ, Cho SN, Kang YH, et al. ^{1}H NMR-based metabolomic profiling in mice infected with *Mycobacterium tuberculosis*. J Proteome Res. 2011;10(5):2238–2247. http://dx.doi.org/10.1021/pr101054m

45. Zhou A, Ni J, Xu Z, Wang Y, Lu S, Sha W, et al. Application of ^{1}H NMR spectroscopy-based metabolomics to sera of tuberculosis patients. J Proteome Res. 2013;12(10):4642–4649. http://dx.doi.org/10.1021/pr4007359

46. Zhou A, Ni J, Xu Z, Wang Y, Zhang W, Wu W, et al. Metabolomics specificity of tuberculosis plasma revealed by ^{1}H NMR spectroscopy. Tuberculosis. 2015;95:294–302. http://dx.doi.org/10.1016/j.tube.2015.02.038

47. Coen M, O'Sullivan M, Bubb WA, Kuchel PW, Sorrell T. Proton nuclear magnetic resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis. Clin Infect Dis. 2005;41(11):1582–1590. http://dx.doi.org/10.1086/497836

48. Qureshi GA, Baig SM, Bednar I, Halawa A, Parvez SH. The neurochemical markers in cerebrospinal fluid to differentiate between aseptic and tuberculous meningitis. Neurochem Int. 1998;32:197–203. http://dx.doi.org/10.1016/S0197-0186(97)00061-2

49. Mason S, Moutloatse GP, Van Furth AM, Solomons R, Van Reenen M, Reinecke CJ, et al. KEMREP: A new qualitative method for the assessment of an analyst's ability to generate a metabolomics data matrix by gas chromatography– mass spectrometry. Curr Metabolomics. 2014;2(1):15–26. http://dx.doi.org/10.2174/2213235X02666140115214427

50. Du Preez I, Loots DT. New sputum metabolite markers implicating adaptations of the host to *Mycobacterium tuberculosis*, and vice versa. Tuberculosis. 2013;93:330–337. http://dx.doi.org/10.1016/j.tube.2013.02.008

51. Halouska S, Chacon O, Fenton RJ, Zinniel DK, Barletta RG, Powers R. Use of NMR metabolomics to analyze the targets of D-cycloserine in mycobacteria: Role of D-alanine racemase. J Proteome Res. 2007;6(12):4608–4614. http://dx.doi.org/10.1021/pr0704332

52. Mason S, Van Furth AM, Mienie LJ, Engelke UFH, Wevers RA, Solomons R, et al. A hypothetical astrocyte–microglia lactate shuttle derived from a ^{1}H NMR metabolomics analysis of cerebrospinal fluid from a cohort of South African children with tuberculous meningitis. Metabolomics. 2015;11:822–837. http://dx.doi.org/10.1007/s11306-014-0741-z

53. Subramanian A, Gupta A, Saxena S, Gupta A, Kumar R, Nigam A, et al. Proton MR CSF analysis and a new software as predictors of the differentiation of meningitis in children. NMR Biomed. 2005;18:213–225. http://dx.doi.org/10.1002/nbm.944

54. Cunha BA. Distinguishing bacterial from viral meningitis: The critical importance of the CSF lactic acid levels. Intensive Care Med. 2006;32(8):1272–1273. http://dx.doi.org/10.1007/s00134-006-0210-x

55. Genton B, Berger JP. Cerebrospinal fluid lactate in 78 cases of adult meningitis. Intensive Care Med. 1990;16(3):196–200. http://dx.doi.org/10.1007/BF01724802

56. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: Acquiring and understanding global metabolite data. Trends Biotechnol. 2004;22:245–252. http://dx.doi.org/10.1016/j.tibtech.2004.03.007

57. Block ML, Hong JS. Microglia and inflammation-mediated neurodegeneration: Multiple triggers with a common mechanism. Prog Neurobiol. 2005;76:77–98. http://dx.doi.org/10.1016/j.pneurobio.2005.06.004

58. Pellerin L, Pellegri G, Bittar PG, Charnay Y, Bouras C, Martin JL, et al. Evidence supporting the existence of an activity-dependent astrocyte-neuron lactate shuttle. Dev Neurosci. 1998;20:291–299. http://dx.doi.org/10.1159/000017324

59. Lee H, Sun E, Ham D, Weissleder R. Chip-NMR biosensor for detection and molecular analysis of cells. Nat Med. 2008;14:869–874. http://dx.doi.org/10.1038/nm.1711

60. El-Kebir M, Van der Kuip M, Van Furth AM, Kirschner DE. Computational modeling of tuberculous meningitis reveals an important role for tumor necrosis factor-α. J Theor Biol. 2013;328:43–53. http://dx.doi.org/10.1016/j.jtbi.2013.03.008

61. De Souza GA, Fortuin S, Aguilar D, Pando RH, McEvoy CR, Van Helden PD, et al. Using a label-free proteomics method to identify differentially abundant proteins in closely related hypo- and hypervirulent clinical *Mycobacterium tuberculosis* Beijing isolates. Mol Cell Proteomics. 2010;9(11):2414–2423. http://dx.doi.org/10.1074/mcp.M900422-MCP200

62. Drennan MB, Nicolle D, Quesniaux VJ, Jacobs M, Allie N, Mpagi J, et al. Toll-like receptor 2-deficient mice succumb to *Mycobacterium tuberculosis* infection. Am J Pathol. 2004;164(1):49–57. http://dx.doi.org/10.1016/S0002-9440(10)63095-7

63. Van Well GTJ, Wieland CW, Florquin S, Roord JJ, Van der Poll T, Van Furth AM. A new murine model to study the pathogenesis of tuberculous meningitis. J Infect Dis. 2007;195:694–697. http://dx.doi.org/10.1086/511273

64. Lee A, Toffaletti DL, Tenor J, Soderblom EJ, Thompson JW, Moseley MA, et al. Survival defects of *Cryptococcus neoformans* mutants exposed to human cerebrospinal fluid result in attenuated virulence in an experimental model of meningitis. Infect Immun. 2010;78(10):4213–4225. http://dx.doi.org/10.1128/IAI.00551-10

65. Wikoff WR, Gangoiti JA, Barshop BA, Siuzdak G. Metabolomics identifies perturbations in human disorders of propionate metabolism. Clin Chem. 2007;53:2169–2176. http://dx.doi.org/10.1373/clinchem.2007.089011

66. Reinecke CJ, Koekemoer G, Van der Westhuizen FH, Louw R, Lindeque JZ, Mienie LJ, et al. Metabolomics of urinary organic acids in respiratory chain deficiencies in children. Metabolomics. 2012;8(2):264–283. http://dx.doi.org/10.1007/s11306-011-0309-0

67. Dercksen M, Duran M, Ijlst L, Mienie LJ, Reinecke CJ, Ruiter JPN, et al. Clinical variability of isovaleric acidemia in a genetically homogeneous population. J Inherit Metab Dis. 2012;35(6):1021–1029. http://dx.doi.org/10.1007/s10545-012-9457-2

A review of the lunar laser ranging technique and contribution of timing systems

AUTHORS:
Cilence Munghemezulu[1,2]
Ludwig Combrinck[1,2]
Joel O. Botai[3]

AFFILIATIONS:
[1]Centre for Geoinformation Science, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa
[2]Space Geodesy Programme, Hartebeesthoek Radio Astronomy Observatory, Krugersdorp, South Africa
[3]South African Weather Service, Pretoria, South Africa

CORRESPONDENCE TO:
Cilence Munghemezulu

EMAIL:
cilence.munghemezulu@up.ac.za

POSTAL ADDRESS:
Centre for Geoinformation Science, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

DATES:
Received: 17 Oct. 2015
Revised: 03 Dec. 2015
Accepted: 04 Dec. 2015

KEYWORDS:
space geodetic techniques; LLR; earth–moon system; retroreflectors

HOW TO CITE:
Munghemezulu C, Combrinck L, Botai JO. A review of the lunar laser ranging technique and contribution of timing systems. S Afr J Sci. 2016;112(3/4), Art. #2015-0400, 9 pages. http://dx.doi.org/10.17159/sajs.2016/20150400

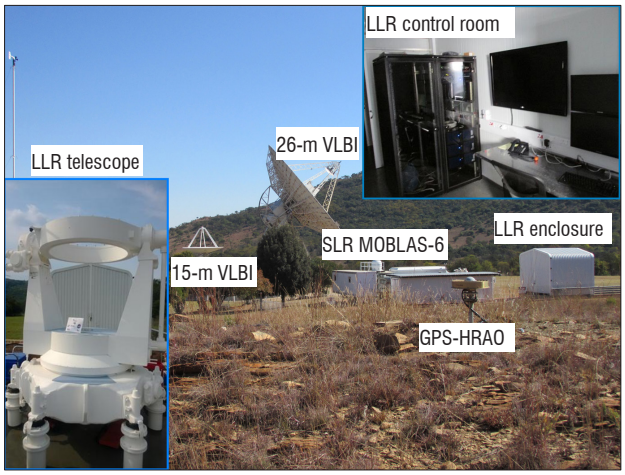
The lunar laser ranging (LLR) technique is based on the two-way time-of-flight of laser pulses from an earth station to the retroreflectors that are located on the surface of the moon. We discuss the ranging technique and contribution of the timing systems and its significance in light of the new LLR station currently under development by the Hartebeesthoek Radio Astronomy Observatory (HartRAO). Firstly, developing the LLR station at HartRAO is an initiative that will improve the current geometrical network of the LLR stations which are presently concentrated in the northern hemisphere. Secondly, data products derived from the LLR experiments – such as accurate lunar orbit, tests of the general relativity theory, earth–moon dynamics, interior structure of the moon, reference frames, and station position and velocities – are important in better understanding the earth–moon system. We highlight factors affecting the measured range such as the effect of earth tides on station position and delays induced by timing systems, as these must be taken into account during the development of the LLR analysis software. HartRAO is collocated with other fundamental space geodetic techniques which makes it a true fiducial geodetic site in the southern hemisphere and a central point for further development of space-based techniques in Africa. Furthermore, the new LLR will complement the existing techniques by providing new niche areas of research both in Africa and internationally.

Introduction

The study of earth's gravity, earth's rotation, geokinematics and inclusion of space-time currently completes the definition of space geodesy.[1] This definition has evolved from the early days of F.R. Helmert, who firstly defined geodesy as the science of the measurement and mapping of the earth's surface.[2,3] It is clear that geodesy as a discipline has two objectives that are closely related: (1) scientific objectives constitute the study of geodynamic phenomena, the gravity field of the earth and other planets, the shape and size of the earth and its orientation in space through earth orientation parameters and (2) practical objectives include cadastral surveying to determine points accurately (up to millimetre level) on the earth's surface, accurate timing, terrestrial geodetic reference frames, and accurate positioning for civil engineering applications.[2] These objectives are realised through space-based techniques: Global Navigation Satellite Systems (GNSS), very long baseline interferometry (VLBI), satellite laser ranging (SLR), lunar laser ranging (LLR) and Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS). Together these techniques are fundamental in defining and maintaining different reference systems and reference frames, and determining three-dimensional station positions, velocities, earth orientation parameters and polar motion with high accuracy, spatial resolution and temporal stability.[4,5]

The accuracy of geodetic products depends on a number of factors such as instrumental accuracies, analysis strategies, accuracy of the models used during data processing (e.g. atmospheric models), third-body perturbation effects and relativistic effects. Among these factors, the stability of timing systems plays a crucial role in determining the accuracy of the measurements. Development of highly stable clocks such as the hydrogen microwave amplification by stimulated emission of radiation (maser) and caesium clocks allows accurate timing and data correlation to be carried out with high levels of accuracy.[6] The application of GNSS in metrology (i.e. time transfer techniques) has advanced the traditional time-keeping services (e.g. mobile reference clocks, terrestrial communication systems using Loran-C or direct radio broadcasts) and allowed comparison of clocks that are distributed around the world with high accuracy and the maintenance of Coordinated Universal Time (UTC).[7] These timing systems contribute towards earth and space observational techniques in ensuring accurate measurements.

Earth observation and space development technologies are very important for the development of a country and have socio-economic benefits such as climatology, land management and monitoring applications. South Africa, Nigeria, Egypt, Algeria, Kenya and, more recently, Ethiopia are African countries that are harnessing space-related technologies ranging from earth observation satellites and space geodesy to radio and optical astronomy.[8,9] The Hartebeesthoek Radio Astronomy Observatory (HartRAO) is located north of Krugersdorp in South Africa (Figure 1). It is collocated with DORIS, GNSS, 26-m and 15-m VLBI telescopes, and an SLR (MOBLAS-6) station. Currently, HartRAO is building a VGOS (VLBI2010) telescope and a new LLR system based on a 1-m aperture telescope donated by the Observatoire de la Côte d' Azur in France.[10] The observatory is a fundamental (fiducial) site in the southern hemisphere as it limits geometrical errors during computation of geodetic or astronomical parameters in the global network and provides high-quality scientific data with high temporal resolution.[10] Here we provide a review of the LLR technique and the importance of timing systems in light of the current development of the new LLR station at HartRAO.



VLBI, very long baseline interferometry; SLR, satellite laser ranging

Figure 1: Geodetic instruments collocated at HartRAO, South Africa. The lunar laser ranging (LLR) azimuth-elevation mount is portrayed in the inset on the left; the tube and other components were removed for upgrade and maintenance. The top right inset depicts the inside of the LLR control room.

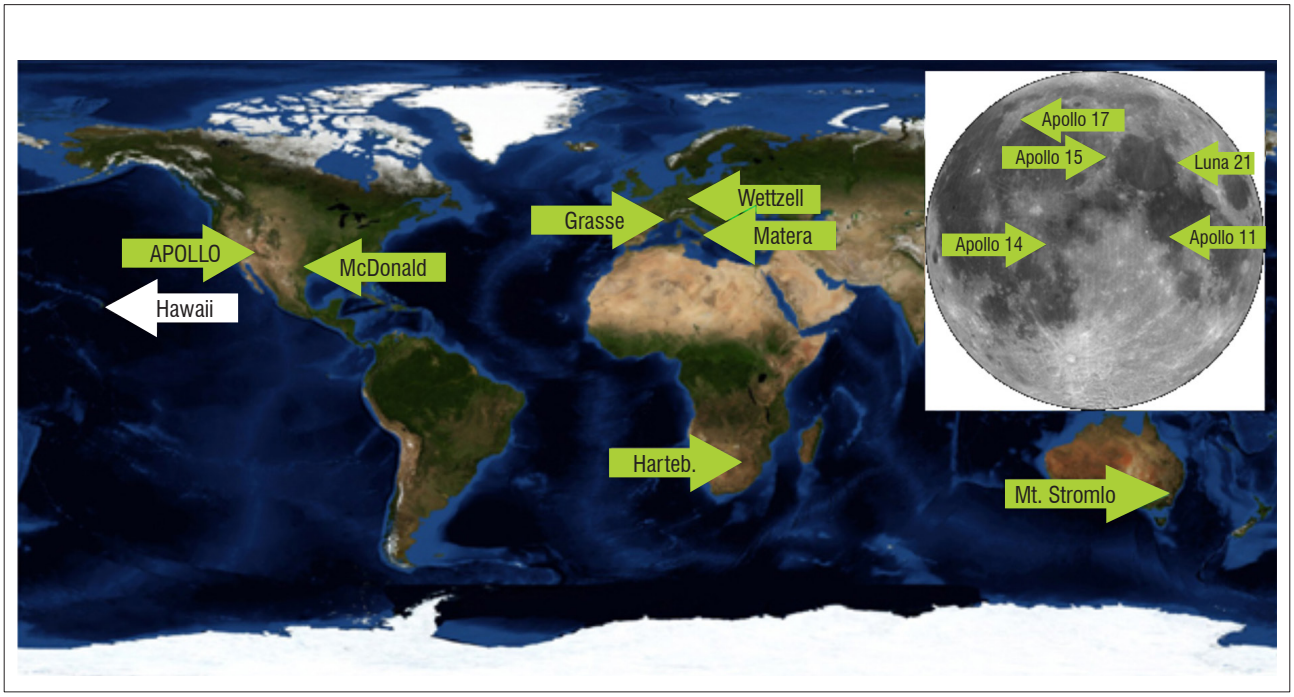
Historical developments of lunar laser ranging

The US Apollo and the Soviet missions deployed arrays of corner-cube retroreflectors on the surface of the moon during the late 1960s and 1970s; the LLR technique has therefore been used to study the earth–moon system since 1969. Currently on the moon, there are Apollo 11, 14 and 15 reflectors deployed by the Apollo missions and Lunokhod I and II reflectors deployed by the Soviet Union[11] (Figure 2). The McDonald Observatory was the first observatory to range to a retroreflector on the moon using a 2.7-m telescope.[12]

Currently there are four LLR stations in the world that range to the moon's retroreflectors on a regular basis. These active stations are (1) the Apache Point Observatory Lunar Laser-ranging Operation (APOLLO), New Mexico, USA; (2) McDonald Laser Ranging Station (MLRS), Texas, USA; (3) the Observatoire de la Côte d' Azur (OCA), France; and (4) Matera, Italy.[13] These stations are located in the northern hemisphere, which results in a weak geometry for the LLR network, as no stations are active in the southern hemisphere. However, the South African radio astronomy observatory (HartRAO), Mount Stromlo SLR observatory in Australia[14], ESO La Silla Observatory IV in Chile[15], National Astronomical Observatory of Japan[16] and the Russians[17] are planning and developing LLR stations. These additional stations will improve and contribute to the current network of the International Laser Ranging Service (ILRS) and to lunar and earth sciences as a whole.

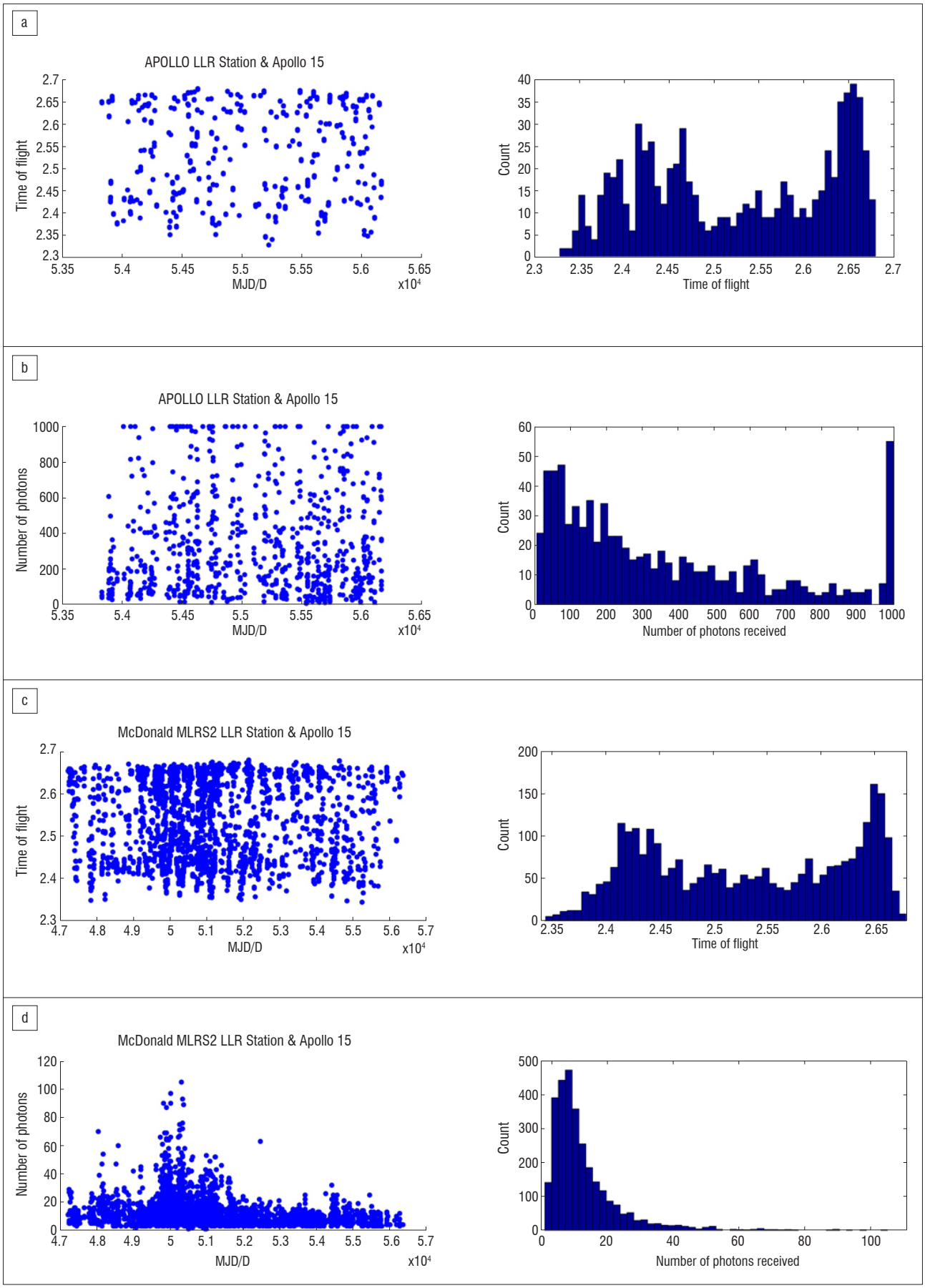
Since the LLR technique was developed, range precision has increased from about 100–250 mm to less than 20 mm.[19] A similar increase in the range precision of the normal points in SLR applications has been reported in Botai et al.[20] and this increase has been attributed to improvements of the internal system components as well as advanced models. System component improvements can be summarised as: (1) replacement of the ruby laser with a Nd:YAG laser, (2) higher emission frequency rates (5–20 Hz), (3) faster and more sensitive detectors and (4) improved timing systems (<20 ns) with lower jitter (<10 ps) resolutions.

The existing LLR stations have provided valuable data to date. In particular, the APOLLO LLR station seems to be the only station that is capable of receiving up to 1000 photons per session (one session typically lasts for 1 h with an expected 2–10 return photons per pulse).[21] In comparison, the photon return rates of the McDonald MLRS2 and Grasse stations range from 1 to 105 per session and from 2 to 605 per session, respectively, when ranging to the Apollo 15 retroreflector on the moon (Figure 3). The data used were provided by the Paris Observatory Lunar Analysis Center (POLAC) and are available to the public (http://polac.obspm.fr/llrdatae.html). Furthermore, these stations indicate that the time-of-flight to the Apollo 15 retroreflector array can range from 2.3 s to 2.7 s.



Source: International Laser Ranging Service[20]

Figure 2: Distribution of the International Laser Ranging Service network of lunar laser ranging stations. The inset map illustrates the location of the retroreflectors on the surface of the moon.



MJD/D, modified Julian day per day

Figure 3: (a,c) Time-of-flight and (b,d) received number of photons measured between the Apollo 15 retroreflector on the moon and the APOLLO and McDonald MLRS2 stations.

Basic principles of lunar laser ranging

Ranging to the moon involves transmitting short laser pulses (about 0.03 m) from the ranging station to the retroreflectors located on the surface of the moon. The two-way time-of-flight of the laser pulses is measured on the ground using highly accurate timing systems. Most SLR stations utilise neodymium-doped yttrium aluminium garnet (Nd:YAG) lasers at a 532-nm wavelength with a repetition rate ranging from 5 Hz to 2 kHz. In SLR applications, laser energies ranging from ~1 to 100 mJ are used to range to artificial satellites such as LAGEOS. Higher laser energies (100–200 mJ) and lower laser pulse frequencies (~20 Hz) as well as larger telescopes are used for LLR applications. These differences are mainly as a result of the greater distance between the moon and the earth compared with distances between satellites that are orbiting the earth.

The basic observational equation is given in Müller et al.[22] as

$$d = c\frac{\tau}{2} = |r_{em} - r_{station} + r_{reflec}| + c\Delta\tau \qquad \text{Equation 1}$$

In Equation 1, d is the station to reflector distance, c is the speed of light in a vacuum, τ is the time-of-flight of the laser pulse, r_{em} describes a vector connecting the geocentre and the selenocentre (centre of mass of the moon), the geocentric position of the observatory is given by $r_{station}$, and r_{reflec} presents the selenocentric position vector of the retroreflector on the moon. The laser pulse is delayed as a result of interaction with the non-linear atmospheric environment, which must be corrected for. Other factors that must be corrected for are discussed later in this paper. The term $\Delta\tau$ describes these corrections, which must be applied to obtain an accurate computed distance.

The earth–moon distance is approximately 385 000 km; unlike in SLR, this distance presents a challenge in LLR because fewer photons are received per single laser shot containing about 10^{15} photons. To illustrate the impact of the earth–moon distance on the photon return rate, Degnan[23] provided a radar link equation that estimates the success of receiving photoelectrons n_{pe} as detected by the receiving telescope as:

$$n_{pe} = \eta_q \left(E_T \frac{\lambda}{hc}\right) \eta_T G_T \sigma \left(\frac{1}{4\pi R^2}\right)^2 A_t \eta_R T_a^2 T_c^2, \qquad \text{Equation 2}$$

where n_q is the detector's quantum efficiency, E_T is the laser pulse energy, λ is the wavelength of the laser, h is Planck's constant, c is the speed of light in a vacuum, n_T is the efficiency of the transmitter optics, G_T is the transmitter gain, σ is the reflector optical cross section and n_R is the efficiency of the receiver optics. The effective area of the telescope receiving aperture is given by A_t. T_a is the one-way atmospheric transmission if present, and T_c is the one-way transmissivity of cirrus cloud. R is the slant range to the moon and can be calculated using:

$$R = -(R_E + h_t)\cos\theta_{zen} + \sqrt{(R_E + h_t)^2\cos^2\theta_{zen} + 2R_E(h_s - h_t) + h_s^2 - h_t^2}, \qquad \text{Equation 3}$$

where R_E is earth's radius (6378 km), h_t is the station height above sea level, h_s is the moon altitude above sea level, and θ_{zen} is the zenith angle of the moon as observed from the ranging station.

The observed raw data (d_{oi}) are filtered to detect gross errors, evaluated to access the accuracy of the observations and compressed for further analysis in the form of a normal point (NP_i) data set (Equation 4).[23] The residuals are computed from the predicted and observed ranges and the outliers are removed using a range window. A suitable trend function (a polynomial fit or a set of orbital parameters) is usually fitted to the residuals to detect further outliers by analysing the deviations from the fit (f_{ri}). The data can then be divided into bins; there is a recommended number of bins to be used for different retroreflectors/satellites, e.g. for LAGEOS 1 and 2, a 120-s bin width is used to divide the data for the duration of the observation. For each bin interval, i, a mean value ($\bar{f}_{ri}$) of all deviations is calculated and added to the trend function at the centre of the interval.[23] This point is referred to as a normal point and thus represents a single observation of the particular interval,

$$NP_i = d_{oi} - (f_{ri} - \bar{f}_{ri}). \qquad \text{Equation 4}$$

The systematic errors, some of which are discussed later, must still be modelled out during further processing. The internal system accuracy of the station (i.e. the root mean square, RMS) can be evaluated using Equation 5 as given in Sinclair[24], where n_i represents the number of observations within the bin.

$$RMS_i = \sqrt{\frac{1}{n_i}\sum_j (f_{ri} - \bar{f}_{ri})^2}. \qquad \text{Equation 5}$$

Timing systems and current progress on the new LLR station

Earth was once thought to function as a perfect clock by counting the number of sunrises or sunsets and constructing a calendar as earth rotates around its axis. Advancements have been made in the more accurate measurement of time and short intervals through the development of mechanical clocks; pendulum clocks were built with an accuracy of about 10 s per day.[25] However, it is important to note that during this period there was no method to synchronise individual clocks.

The development of timing systems has also led to the development of many technologies that are of societal benefit and that we take for granted, such as GPS, the electric power grid and mobile phones.[26] The reported accuracy of the first operational atomic clock was 0.1 ms/day.[27] Accuracy of caesium atomic clocks increased with the development of laser cooling technologies; the National Institute of Standards and Technology (NIST) has so far developed NIST-F1 and NIST-F2, which have an accuracy of more than 10^{-15} at 1 s – the most accurate caesium clocks to date.[28] Quartz and rubidium clocks are used in scientific applications such as LLR and SLR as they are relatively affordable and also offer a high accuracy of about 10^{-12} at 1 s.

HartRAO currently is developing a LLR station by utilising a 1-m aperture Cassegrain telescope donated by the French Observatoire de la Côte d' Azur.[29] The LLR station is designed to range to retroreflectors mounted on satellites and the surface of the moon. The LLR station will use a newly developed 4393A rubidium timing reference system by Microsemi with an accuracy to sub-picosecond ($<10^{-12}$) level. This system will improve the measurement of time-of-flight of the laser pulses and limit instrumentation error dependency. Detailed current development and future perspectives of the LLR system are described in Combrinck[10].

Factors contributing to range bias

In light of the new LLR analysis software being developed at HartRAO, it is necessary to take into account all the factors that affect ranging to the moon and satellites, and which eventually affect the accuracy of determining the moon's orbit. In terms of SLR, these factors are listed in Combrinck and Suberlak[30] and Combrinck[31] as: earth's geopotential[32], solid earth tides[33], ocean tides, planetary third-body perturbation (sun, moon and planets), relativistic acceleration[34], atmospheric tide and atmospheric drag[35], solar radiation pressure[36], earth radiation pressure, thermal radiation acceleration[37], lunar librations[38] (for LLR), Shapiro delay[39], tropospheric delay[40] and delay induced by electronic systems. Two separate software suites will be used for either SLR or LLR analysis, because although some of the corrections are similar (e.g. station displacement due to solid earth tides), the analysis problem is quite different. Not all of these factors are discussed here; further information can be found in the references provided. Corrections for a few factors are applied to illustrate the importance of considering these factors in the LLR analysis package. In a simplified version, the time-of-flight of the laser pulses can be described by

$$ToF = T_{sy} + D_{at} + G_{rt} + T_{td} + \varepsilon \qquad \text{Equation 6}$$

where T_{sy} represents actual time interval measurement by the station, D_{at} is the time delay due to the atmosphere, G_{rt} is the general relativity correction, T_{td} is the time variation induced by tidal effects and ε includes all other corrections not listed above and unknowns. Time-of-flight

is the actual time interval measurement at the station and comprises several components. The station timing system measures the gross sum of contributions as a time interval. T_{sy} is the 'true' satellite distance (expressed as time-of-flight). Data processing is applied mainly to account for the true satellite distance by removing the other disturbing contributions. Practical examples are given below for selected parameters to illustrate their importance.

Tidal correction on station position

The gravitational attractions on earth of the sun, moon and planets, result in a force that deforms the earth's gravitational field and induces solid earth tides.[41] This force is coupled with ocean and atmospheric loading effects as well as the mantle convection processes within the earth. The earth system responds to these effects through mass displacement, rotational acceleration and continuous deformation of the solid crust.[30] Space geodetic instruments such as LLR are affected by these continuous deformation effects, which translate to an additional range bias during ranging. Hence, the LLR analysis software (currently under development at HartRAO) must be able to model these effects with high accuracy in order to improve the range bias. The earth tide, pole tide and ocean tide effects are well described in McCarthy and Petit[42] and readers are referred to this reference for more information.

The solid earth tides can be conveniently modelled[43] as variations in the standard geopotential coefficients C_{nn} and S_{nm}, and can be described as

$$\Delta\bar{C}_{nm} - i\Delta\bar{S}_{nm} = \frac{k_{nm}}{2n-1}\sum_{j=2}^{3}\frac{GM_j}{GM_E}\left(\frac{R_e}{r_j}\right)^{n+1}\bar{P}_{nm}(\sin\Phi_j)e^{-im\lambda j}\{with\ \bar{S}_{n0}=0\}$$

Equation 7

where k_{nm} is the nominal Love number for degree n and order m, R_e represents the equatorial radius of the earth, GM_E is the gravitational parameter for the earth; the gravitational parameter for the moon ($j=2$) and sun ($j=3$) is given by GM_j, the distance to the geocentre of the moon or sun is represented by r_j, Φ_j is the body fixed geocentric latitude of the moon or sun, λ_j represents body fixed east longitude (from Greenwich) of the moon or sun and $\bar{P}_{nm}$ represents the normalised associated Legendre function.

A library developed in Fortran to compute station displacement due to earth tides was developed by Petrov[33] and is currently used by the Satellite Data Analysis Software (SDAS) developed at HartRAO[30]. The same library was also used in this study to illustrate the effects of earth tide on the earth crust for the HartRAO site, where the LLR telescope is located. A continuous station displacement can be clearly seen, with magnitudes ranging from -160 mm to 300 mm in vertical displacement (Figure 4).

Gravitational pull by the sun, moon and planets also results in ocean tides; an additional weight by ocean loading influences crustal displacement and results in temporal variations of station position. Stations that are inland are expected to be less affected than those that are close to the coast. There are plans to move the new LLR station at HartRAO, once completed, to Matjiesfontein in the Great Karoo (in the Western Cape of South Africa).[29] This new site is about 240 km from the Southern Ocean; the effect of ocean tides is expected to be greater at the new site than at the current location (HartRAO, Krugersdorp). In order to better understand and model the effects of ocean and earth tides, atmospheric loading and local hydrological cycles, instruments such as a superconducting gravimeter are required, and should be mounted as close as possible to the ranging telescope to measure small displacements. A typical example is the APOLLO station at Apache Point, Texas: their gravimeter has a noise level of 1 nm/s^2, which is very sensitive and can model these effects to millimetre accuracy or better.[21] HartRAO has installed a gravimeter to compensate for these effects and improve the existing models.

Tropospheric delay correction

The troposphere introduces a significant delay (up to several metres) both in radio and optical wavelengths. Different models have been developed to account for this delay in optical wavelength observations[44] and radio wavelength observations[40]. The delay increases with decreasing elevation angles. Usually, observations are made at higher elevation angles (e.g. above 20° for SLR or LLR). Data from a network of radiosonde distributed worldwide are used to create and validate tropospheric models. A series of mapping functions is used to account for tropospheric delay in the form of[40]:

$$\Delta L(\varepsilon) = \Delta L_h^z \cdot mf_h(\varepsilon) + \Delta L_w^z \cdot mf_w(\varepsilon),$$

Equation 8

where the total slant delay ΔL as a function of an elevation angle ε is expressed as the sum of a hydrostatic and a wet portion – both can be expressed as the product of a zenith delay and a corresponding mapping function. To illustrate the extent of the delay introduced by the troposphere at different elevation angles for radio frequencies, we used the GPT2w model developed by Böhm et al.[40] to estimate slant delay for the HartRAO site. Figure 5 depicts computed slant delay correction at different elevations. An example of the delay in optical wavelength (532 nm) for a 7-day arc of several SLR stations is given in Combrinck[31], based on the model developed by Mendes and Pavlis[44]. It must be noted that the delay in radio and optical wavelengths is different but the pattern is the same (i.e. the delay increases with a decrease in elevation angle). For the purpose of LLR analysis software, appropriate models will be implemented to compensate for atmospheric effects.

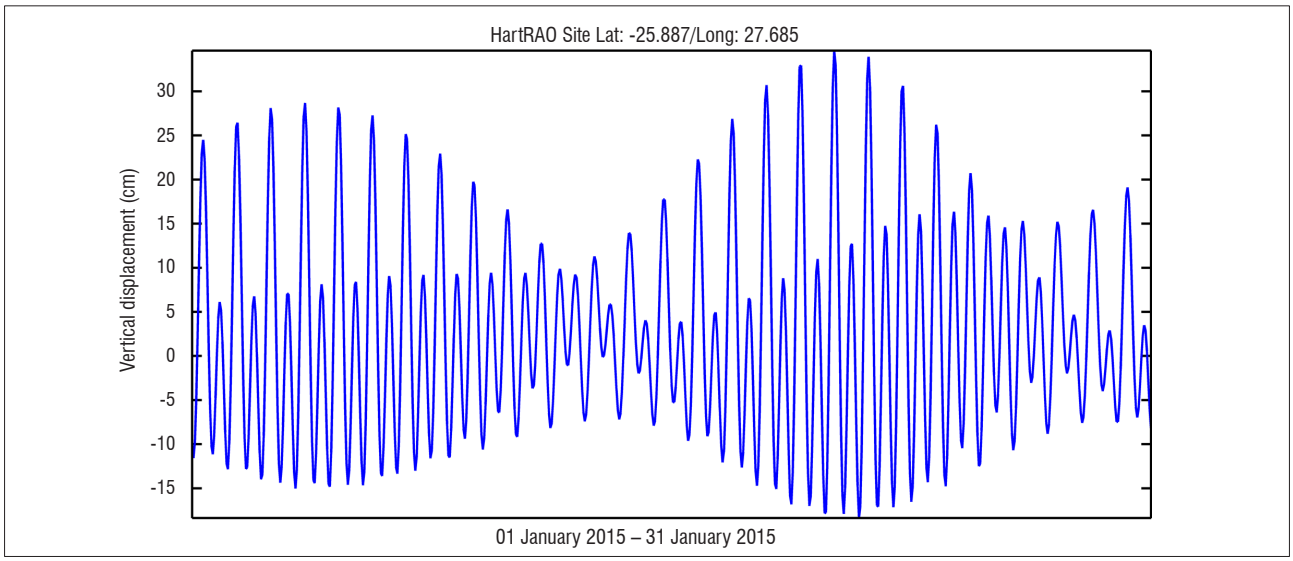


Figure 4: Station vertical displacement for the HartRAO site, simulated for the month of January 2015.

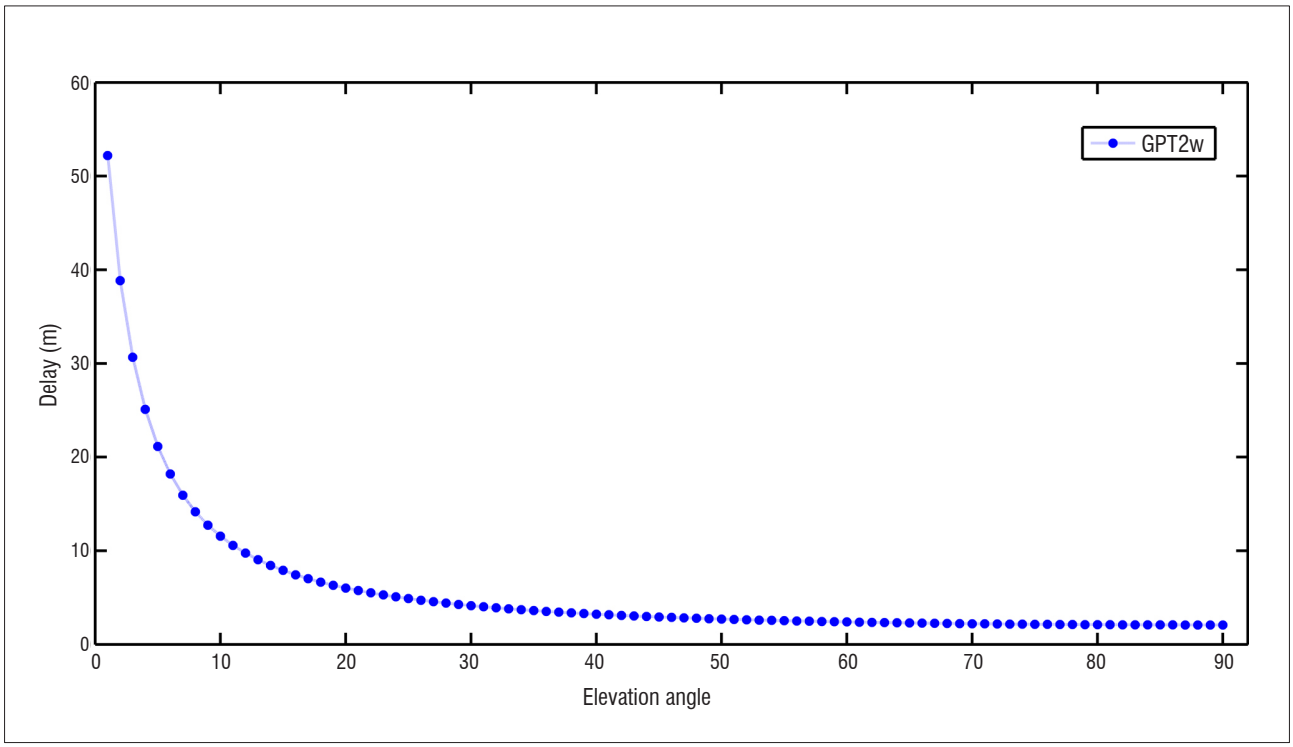


Figure 5: Slant delay as a function of elevation angle for the HartRAO site, computed for the day of 30 June 2015, using Böhm et al.'s[40] GPT2w model.

Timing systems

Timing systems are crucial in space geodetic and astronomical instrumentation. In particular, the accuracy of measurements depends on the accuracy of the reference timing systems. Most reference timing stations for SLR and LLR are at the level of 10^{-12} Allan deviation at 1 s using rubidium/quartz crystal oscillators[45], while the hydrogen maser clocks used in VLBI techniques are better than 10^{-14} at 1 s.[46] Most stations are limited in their use of highly accurate timing systems such as maser clocks because of their high cost. The noise levels within event timers and photon detection systems also play a crucial role in the accuracy of the normal data points and range bias. It is envisaged that the new LLR station will use the newly developed 4380A-GPS disciplined master timing reference by Microsemi Corporation. This unit has an Allan deviation of 10^{-13} at 1 s and less than 10 ns RMS timing accuracy. A low jitter event timer of 3 ps RMS and a solid state photon detector (a single-photon avalanche diode or SPAD) with a quantum efficiency of 50% will be integrated to allow for sub-centimetre ranging precision.

An impact of timing systems on the observed range measurements can be clearly seen from Figure 6, in which a normal point data quality of the Potsdam SLR station between 2003 and 2011 indicates high variation, ranging from 10 mm to 25 mm.[45] During 2011, an old timing system was replaced with a more modern timing system with better specifications (Table 1). The SPAD with a quantum efficiency of 28% was replaced with a SPAD with a quantum efficiency of 40% and low jitter. As a result, the normal point data are characterised by low variations, which can be directly linked to the improvement of the timing systems (among others) post 2011.

Lunar laser ranging contributions to science and society

Earth observation techniques have a direct impact on the social lives of people. Imagery from remote sensing techniques is widely applied in areas such as agriculture[47], ecosystem management[48,49], water management[50], disaster management[51] and weather applications[52]. These applications provide examples of the use of earth observation techniques to assist in planning, early warning systems for natural disasters, and management of earth's resources. More advanced remote sensing techniques such as VLBI, GNSS, LLR and SLR contribute in the same way as satellite remote sensing does to society. The derived data products – such as the International Terrestrial Reference Frame (ITRF), the International Celestial Reference Frame, earth orientation parameters, the gravity field, and atmosphere and ionosphere parameters – form part of the foundation of earth observation technologies.[53] A set of station coordinates and velocities derived from the geodetic techniques is used to construct a reference frame that allows connection and comparison between different geodetic data sets over varying space and time. This construction is done through combination of the data sets using scientific software such as CATREF[54] and by taking into account the local site ties. The latest ITRF2008 has an accuracy to sub-centimetre.[4] This system provides a basis for local reference frame systems, which are realised based on the ITRF, including, for example, the unified African Geodetic Reference Frame (AFREF), which can be used for cadastral surveys, mapping and civil engineering applications.[55]

The LLR technique in particular contributes towards advancement of the field of space geodesy, lunar science, earth–moon system dynamics and gravitational physics. The increased accuracy in range measurements from 200–300 mm in early development stages to about 20–30 mm in recent developments, has provided ways to test and evaluate general relativity theory[56] and the gravitational constant with ranging accuracy at picosecond level. Williams et al.[57] derived geophysical and orbit parameters of the moon; the gravitational constant was evaluated to be $\dot{G}/G = 4 \pm 9 \times 10^{-13}$ per year by Williams et al.[58] and a more recent value of $-0.7 \pm 3.8 \times 10^{-13}$ per year is reported by Hofmann et al.[59] The LLR system at HartRAO is being developed by HartRAO staff and PhD and MSc students registered at various South African universities, hence capacity building and skills transfer are at a high level. This project will support environmental monitoring through proxy parameters which measure, for example, the state of the atmosphere, gravity fields (for groundwater storage monitoring) and seismic activities, which are important to society.

Table 1: Historical improvement of the Potsdam satellite laser ranging station timing system[45]

Timing equipment	Date of installation/replacement		
	20 July 2001	**2001–2004**	**01 May 2011**
Detection type	SPAD (AD230)		SPAD (MPD-ICTC)
Quantum efficiency (%)	28		40
Jitter (ps)	75		20
Signal processor	CFD (TC4S4)		
Time measurement	Interval	Event	Event
Model	SR620	A031-ET	A032-ET
Resolution (ps)	4	1	1
Precision (ps)	20	10	7

SPAD, single-photon avalanch diode; CFD, constant fraction discriminator

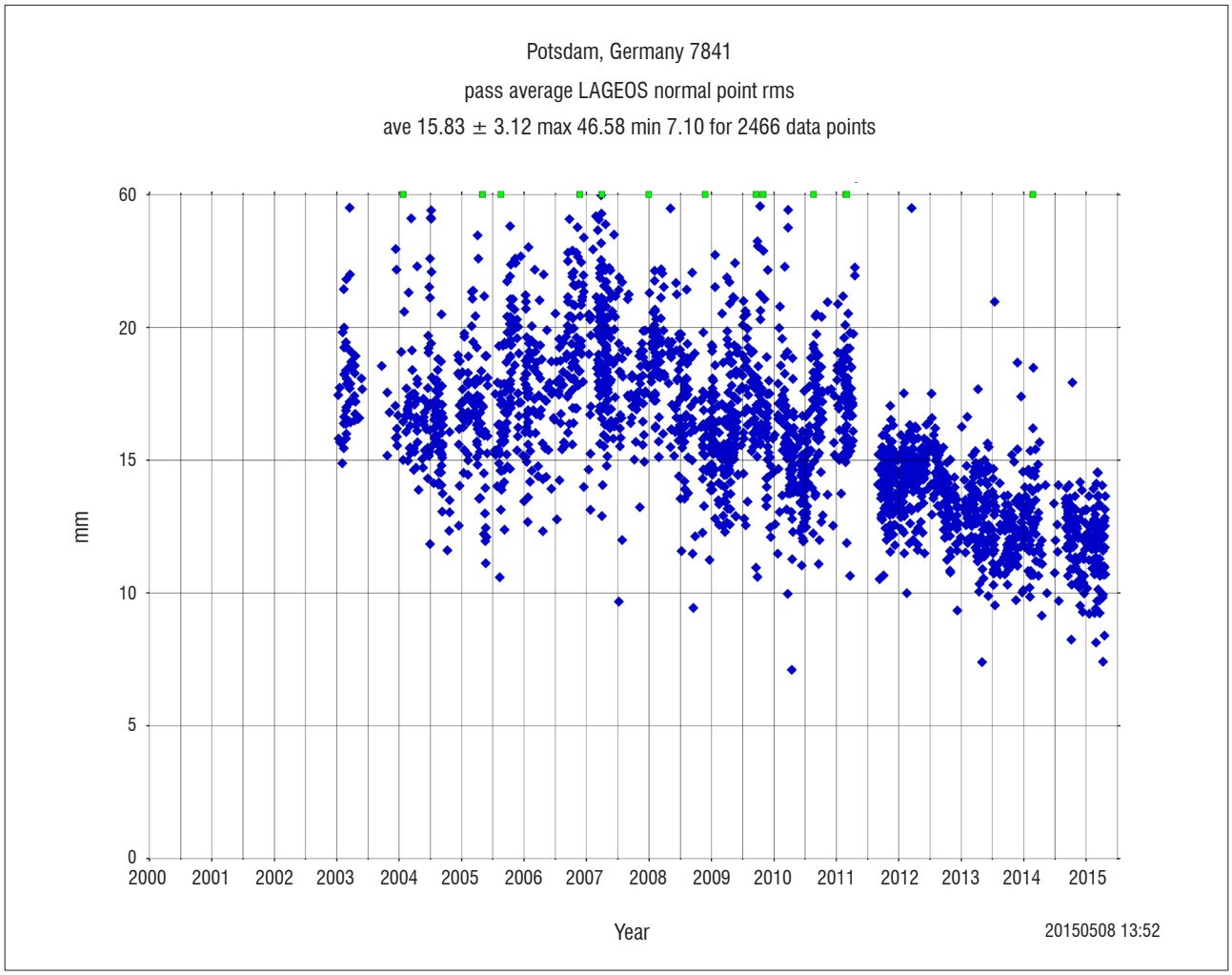


Figure 6: Potsdam satellite laser ranging station performance, measured based on normal point root mean square for LAGEOS ranging data. The single-photon avalanche diode was replaced with one of higher specifications in 2011.[45]

Conclusion

We have briefly reviewed the LLR technique and the timing systems in light of the new LLR station at HartRAO. The difficulties associated with ranging to the moon were highlighted, and factors that contribute to range bias were emphasised as these must be incorporated in the LLR analysis software. This new LLR station will contribute to both local and global communities to meet the scientific objectives of the currently growing space science endeavours by many countries as well as support socio-economic developments. Existing LLR stations are sparsely distributed globally and the station at HartRAO (currently in development) together with Mount Stromlo SLR observatory in Australia have the opportunity to expand the existing global network to the southern hemisphere.

Data products derived from highly technical space geodetic techniques such as LLR have indirect and direct benefits to society, hence the LLR project at HartRAO has received local support and international support from organisations such as the National Aeronautics and Space Administration (USA) and the Observatoire de la Côte d' Azur (France). There are a number of factors that must be taken into consideration during the implementation of the LLR analysis software. The first step is to ensure that highly accurate (to sub-picosecond level) timing sub-systems are implemented to minimise local systematic errors. The delay induced by the environment can be modelled with current existing algorithms to a high level of confidence. This new LLR station will open many opportunities for current and future space programmes, with societal benefits, both in Africa and internationally.

Acknowledgements

This research was financially supported by the National Research Foundation (NRF), the Department of Science and Technology and Inkaba yeAfrica. Lunar laser ranging data were provided by the Paris Observatory Lunar Analysis Center (POLAC). Support from Observatoire de la Côte d' Azur (OCA) and NASA is also acknowledged.

Authors' contributions

C.M. conceptualised, designed and wrote the original draft manuscript; L.C. and J.O.B. modified, edited and approved the manuscript.

References

1. Combrinck L. A comparison of general relativity theory evaluations using VLBI and SLR: Will GGOS improve these results? In: Behrend D, Baver K, editors. IVS 2012 General Meeting Proceedings; 2012 March 4–9; Madrid, Spain. Greenbelt, MD: International VLBI Service for Geodesy and Astrometry; 2012. p. 357–361.
2. Torge W, Müller J. Geodesy. 4th ed. Berlin: Walter de Gruyter; 2012. p. 433. http://dx.doi.org/10.1515/9783110250008
3. Helmert FR. Die mathematischen und physikalischen Theorieen der höheren Geodäsie [The mathematical and physical theories of higher geodesy]. Leipzig: Die mathematischen Theorieen; 1880.
4. Altamimi Z, Collilieux X, Metivier L. ITRF2008: An improved solution of the international terrestrial reference frame. J Geod. 2011;85(5):457–473. http://dx.doi.org/10.1007/s00190-011-0444-4
5. Lu Z, Qu Y, Qiau S. Geodesy: Introduction to geodetic datum and geodetic systems. Berlin: Springer-Verlag; 2014. http://dx.doi.org/10.1007/978-3-642-41245-5
6. Wynands R, Weyers S. Atomic fountain clocks. Metrologia. 2005;42(3):S64. http://dx.doi.org/10.1088/0026-1394/42/3/S08
7. Lombardi MA. Chapter 17: Fundamentals of time and frequency. In: Bishop RH, editor. The mechatronics handbook. Boca Raton, FL: CRC Press; 2002. Available from: http://www.sze.hu/~szenasy/Szenzorok%20%E9s%20aktu%E1torok/Szenzakt%20jegyzetek/Mechatronics%2520handbook%5B1%5D.pdf
8. Martinez P. Space science and technology in South Africa: An overview. African Skies/Cieux Africains. 2008;12:46–49.
9. Ngcofe L, Gottschalk K. The growth of space science in African countries for earth observation in the 21st century. S Afr J Sci. 2013;109(1/2), Art. #a001, 5 pages. http://dx.doi.org/10.1590/sajs.2013/a001
10. Combrinck L. Development of a satellite and lunar laser ranger and its future applications. Paper presented at: The 62nd International Astronautical Congress; 2011 October 03–07; Cape Town, South Africa. Paris: International Astronautical Federation; 2011. IAC-11-A2.1. http://dx.doi.org/10.13140/2.1.1743.3928
11. Williams JG, Newhall XX, Dickey JO. Relativity parameters determined from lunar laser ranging. Phys Rev D. 1996;53(12):6730–6738. http://dx.doi.org/10.1103/PhysRevD.53.6730
12. Barker ES, Calame O, Mulholland JD, Shelus PJ. Improved coordinates for Lunokhod 2 based on laser observations from McDonald Observatory. Space Res. 1975;XV:71–74.
13. Hofmann F, Müller J, Biskupek L, Mai E, Torre JM. Lunar laser ranging – What is it good for? Paper presented at: The 18th International Workshop on Laser Ranging; 2013 November 11–15; Fujiyoshida, Japan. Available from: http://cddis.gsfc.nasa.gov/lw18/docs/papers/Session9/13-04-02-MuellerJM.pdf
14. Greene B, McK Luck J. LLR developments at Mount Stromlo. Paper presented at: The 13th International Workshop on Laser Ranging; 2002 October 07–11; Washington DC, USA. Available from: http://cddis.gsfc.nasa.gov/lw13/docs/papers/llr_greene_1m.pdf
15. Fienga A, Courde C, Torre JM, Manche H, Murphy T, Mueller J, et al. Interests of a new lunar laser instrumentation on the ESO NTT telescope [arXiv:1405.0473 astro-ph.IM]. c2014 [cited 2015 Jun 15]. Available from: http://arxiv.org/abs/1405.0473
16. Noda H, Kunimori H, Araki H. Lunar laser ranging experiment at Koganei SLR station. Presented at: The 45th Lunar and Planetary Science Conference; 2014 March 17–21; The Woodlands, TX, USA. Available from: http://www.hou.usra.edu/meetings/lpsc2014/pdf/1638.pdf
17. Vasilyev MV, Yagudina EI, Torre J-M, Feraudy D. Planned LLR station in Russia and its impact on the lunar ephemeris accuracy [article on the Internet]. c2014 [cited 2015 Sep 01]. Available from: http://syrte.obspm.fr/jsr/journees2014/pdf/Vasilyev.pdf
18. International Laser Ranging Services (ILRS). LLR map of stations [image on the Internet]. c2009 [updated 2015 Feb 04; cited 2015 May 17]. Available from: http://ilrs.gsfc.nasa.gov/science/scienceContributions/lunar.html
19. Murphy T. Lunar laser ranging: A laboratory for gravity. Proceedings of SLAC Conferences, Workshops & Symposiums, 2007/08/09 [document on the Internet]. c2007 [cited 2015 Jun 30]. Available from: http://www-conf.slac.stanford.edu
20. Botai CM, Combrinck L, Botai JO. Satellite laser ranging measurements in South Africa: Contributions to earth system science. S Afr J Sci. 2015;111(3/4), Art. #2013-0193, 9 pages. http://dx.doi.org/10.17159/sajs.2015/20130193
21. Murphy T, Adelberger E, Battat J, Colmenares N, Crossley D, Holye CD, et al. APOLLO performance and data quality. Paper presented at: The 19th International Workshop on Laser Ranging; 2014 October 27–31; Annapolis, MD, USA. Available from: http://cddis.gsfc.nasa.gov/lw19/docs/2014/Papers/3061_Murphy_paper.pdf
22. Müller J, Williams JG, Turyshev SG. Lunar laser ranging contributions to relativity and geodesy. In: Dittus H, Lammerzahl C, Turyshev SG, editors. Laser, clocks and drag-free control: Exploration of relativistic gravity in space. Berlin: Springer; 2008. p. 457–472. http://dx.doi.org/10.1007/978-3-540-34377-6_21
23. Degnan J. Millimetre accuracy satellite laser ranging. In: Smith DE, Turcotte DL, editors. Contributions of space geodesy to geodynamics: Crustal dynamics. Geodynamics Series 25. Washington DC: American Geophysical Union; 1993. p. 133–162.
24. Sinclair AT. Data screening and normal point formation: Re-Statement of Herstmonceux Normal Point Recommendation [homepage on the Internet]. c1997 [updated 2015 Feb 04; cited 2015 May 17]. Available from: http://ilrs.gsfc.nasa.gov/data_and_products/data/npt/npt_algorithm.html
25. Lombardi MA. The evolution of time measurement part 2: Quartz clocks. IEEE Instrumentation & Measurement Magazine. 2011;14(5):41–48.

26. Lombardi MA. The evolution of time measurement part 3: Atomic clocks. IEEE Instrumentation & Measurement Magazine. 2011;14(6):46–49. http://dx.doi.org/10.1109/MIM.2011.6086901

27. Henderson D. Essen and the National Physical Laboratory's atomic clock. Metrologia. 2005;42(3):S4. http://dx.doi.org/10.1088/0026-1394/42/3/S02

28. Heavner TP, Donley EA, Levi F, Costanzo G, Parker TE, Shirley JH, et al. First accuracy evaluation of NIST-F2. Metrologia. 2014;51(3):174–184. http://dx.doi.org/10.1088/0026-1394/51/3/174

29. Combrinck L, Botha R. Challenges and progress with the development of a lunar laser ranger for South Africa. Paper presented at: The 18th International Workshop on Laser Ranging; 2013 November 11–15; Fujiyoshida, Japan. Available from: http://cddis.gsfc.nasa.gov/lw18/docs/papers/Session13/13-05-04-Combrinck.pdf

30. Combrinck L, Suberlak V. Earth-tide as parameter of crustal motion correction for SLR station displacement. S Afr J Geol. 2007;110(2–3):203–210. http://dx.doi.org/10.2113/gssajg.110.2-3.203

31. Combrinck L. Satellite laser ranging. In: Xu G, editor. Sciences of geodesy I: Advances and future directions. Berlin: Springer-Verlag; 2010. p. 302–336. http://dx.doi.org/10.1007/978-3-642-11741-1_9

32. Botai MC, Combrinck L. Investigating the accuracy of gravity field models using satellite laser ranging data. S Afr J Geol. 2011;114(3–4):539–544. http://dx.doi.org/10.2113/gssajg.114.3-4.535

33. Petrov L. Software sotid for computation of site displacements due to the solid earth tides: Updated pdf documentation 2005.02.11 [document on the Internet]. c2005 [cited 2015 Jun 30]. Available from: http://gemini.gsfc.nasa.gov/sotid

34. Combrinck L. General relativity and space geodesy. In: Xu G, editor. Sciences of geodesy II: Innovation and future developments. Berlin: Springer-Verlag; 2013. p. 53–95. http://dx.doi.org/10.1007/978-3-642-28000-9_2

35. Van Dam TM, Herring TA. Detection of atmospheric pressure loading using very long baseline interferometry measurements. J Geophys Res. 1994;99(B3):4505–4517. http://dx.doi.org/10.1029/93JB02758

36. Ziebart M. High precision analytical solar radiation pressure modelling for GNSS spacecraft [PhD thesis]. London: University of East London; 2001.

37. Vokrouhlický D. A note on the solar radiation perturbations of lunar motion. Icarus. 1997;126(2):293–300. http://dx.doi.org/10.1006/icar.1996.5652

38. Mulholland JD, Silverberg EC. Measurement of physical librations using laser retroreflectors. Moon. 1972;4(1–2):155–159. http://dx.doi.org/10.1007/BF00562923

39. Shapiro LI. Fourth test of general relativity. Phys Rev Lett. 1964;13:789–791. http://dx.doi.org/10.1103/PhysRevLett.13.789

40. Böhm J, Möller G, Schindelegger M, Pain G, Weber R. Development of an improved empirical model for slant delays in the troposphere (GPT2w). GPS Solutions. 2015;19(3):433–441. http://dx.doi.org/10.1007/s10291-014-0403-7

41. Métivier L, Viron O, Conrad CP, Renault S, Diament M, Patau G. Evidence of earthquake triggering by the solid earth tides. Earth Planet Sci Lett. 2009;278(3–4):370–375. http://dx.doi.org/10.1016/j.epsl.2008.12.024

42. McCarthy D, Petit G, editors. IERS conventions (2003). IERS Technical Note 32. Frankfurt: Verlag des Bundesamts fr Kartographie und Geodsie; 2004. Available from: http://www.iers.org/SharedDocs/Publikationen/EN/IERS/Publications/tn/TechnNote32/tn32.pdf?__blob=publicationFile

43. Eanes RJ, Schutz B, Tapley B. Earth and ocean tide effects on Lageos and Starlette. In: Kuo JT, editor. Proceedings of the Ninth International Symposium on Earth Tides. Stuttgart: Sckweizerbart'sche Verlagabuchhandlung; 1983. p. 239–249.

44. Mendes VB, Pavlis EC. High-accuracy zenith delay prediction at optical wavelengths. Geophys Res Lett. 2004;31(14):L14602. http://dx.doi.org/10.1029/2004GL020308

45. International Laser Ranging Services (ILRS). LLR station log files [image on the Internet]. c2009 [updated 2015 Feb 04; cited 2015 Jun 30]. Available from: http://ilrs.gsfc.nasa.gov/network/stations/index.html

46. International VLBI Service for Geodesy and Astrometry. Station log files [homepage on the Internet]. c2015 [updated 2015 Jun 30; cited 2015 Jun 30]. Available from: http://ivscc.gsfc.nasa.gov/about/org/components/ns-list.html

47. Bastiaanssen WGM, Molden DJ, Makin LW. Remote sensing for irrigated agriculture: Examples from research and possible applications. Agr Water Manage. 2000;46(2):137–155. http://dx.doi.org/10.1016/S0378-3774(00)00080-9

48. Kerr TJ, Ostrovsky M. From space to species: Ecological applications for remote sensing. Trends Ecol Evol. 2003;18(6):299–305. http://dx.doi.org/10.1016/S0169-5347(03)00071-5

49. Cohen WB, Goward SN. Landsat's role in ecological applications of remote sensing. BioScience. 2004;54(6):535–545. http://dx.doi.org/10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2

50. Jackson TJ, Schmugge J, Engman ET. Remote sensing applications to hydrology: Soil moisture. Hydrolog Sci J. 1996;41(4):517–530. http://dx.doi.org/10.1080/02626669609491523

51. Madry S. Space systems for disaster warning, response, and recovery. New York: Springer; 2015. http://dx.doi.org/10.1007/978-1-4939-1513-2

52. Lazzara MA, Coletti A, Diedrich BJ. The possibilities of polar meteorology, environmental remote sensing, communications and space weather applications from Artificial Lagrange Orbit. Adv Space Res. 2011;48(11):1880–1889. http://dx.doi.org/10.1016/j.asr.2011.04.026

53. Rothacher M. Towards a rigorous combination of space-geodetic techniques. In: Richter B, Schwegmann W, Dick WR, editors. Proceedings of the IERS Workshop on Combination Research and Global Geophysical Fluids; 2002 November 18–21; Munich, Germany. Frankfurt: Verlag des Bundesamts für Kartographie und Geodäsie; 2003. p. 7–18.

54. Altamimi Z, Sillard P, Boucher C. CATREF software: Combination and analysis of terrestrial reference frames. Technical manual [document on the Internet]. c2006 [cited 2015 Jul 02]. Available from: http://grgs.obs-mip.fr/en/content/download/303/2351/file/CATREF-1.pdf

55. Drewes H, Hornik H. Travaux volume 38: Reports 2011–2013. Munich: International Association of Geodesy; 2013. Available from: http://iag.dgfi.tum.de/fileadmin/IAG-docs/Travaux_2011-2013.pdf

56. Müller J, Williams J, Turyshev S, Shelus P. Potential capabilities of lunar laser ranging for geodesy and relativity. In: Tregoning P, Rizos C, editors. Dynamic planet. IAG Symposia volume 130. Berlin: Springer; 2007. p. 903–909. http://dx.doi.org/10.1007/978-3-540-49350-1_126

57. Williams JG, Boggs DH, Folkner WM. DE421 lunar orbit, physical librations, and surface coordinates [Interoffice memorandum JPL IOM 335-JW, DB, WF-20080314-001]. 2008 March 14. Available from: ftp://ssd.jpl.nasa.gov/pub/eph/planets/ioms/de421_moon_coord_iom.pdf

58. Williams JG, Turyshev SG, Boggs DH. Progress in lunar laser ranging tests of relativistic gravity. Phys Rev Lett. 2004;93(26), Art. #261101, 4 pages. http://dx.doi.org/10.1103/PhysRevLett.93.261101

59. Hofmann F, Müller J, Biskupek L. Lunar laser ranging test of the Nordtvedt parameter and a possible variation in the gravitational constant. A&A. 2010;522, Art. #L5, 3 pages. http://dx.doi.org/10.1051/0004-6361/201015659

AutoCal: A software application for calibrating photometric data

AUTHORS:
Daniël J. Wium[1]
Brian van Soelen[2]

AFFILIATIONS:
[1]Department of Computer Science and Informatics, University of the Free State, Bloemfontein, South Africa
[2]Department of Physics, University of the Free State, Bloemfontein, South Africa

CORRESPONDENCE TO:
Daniël Wium

EMAIL:
WiumDJ@ufs.ac.za

POSTAL ADDRESS:
Department of Computer Science and Informatics, University of the Free State, PO Box 339, Bloemfontein 9300, South Africa

DATES:
Received: 27 Jan. 2015
Revised: 02 June 2015
Accepted: 24 Aug. 2015

KEYWORDS:
astrophysics; software; photometry; Be/X-ray binaries; calibration

HOW TO CITE:
Wium DJ, Van Soelen B. AutoCal: A software application for calibrating photometric data. S Afr J Sci. 2016;112(3/4), Art. #2015-0034, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20150034

We present a software application for the calibration of stellar magnitudes in the absence of standard stars. It uses an existing algorithm to match stars in the target's field of view to catalogue entries and computes the average offset between the two sets of magnitudes using a weighted least-squares approach. This offset is used to calibrate the target's instrumental magnitude. The software application was used to calibrate magnitudes for six Be/X-ray binary sources in the Small Magellanic Cloud and the results were compared with published results for these sources. Where comparisons were possible, our results agreed with those results within the uncertainties specified. Infrared variability was found for all six of the sources tested. The interactive outlier removal that was made possible by our software allowed for smaller uncertainties to be reported for our results.

Introduction

The measured (instrumental) magnitude of a star, a measure of its brightness, is influenced by loss of light in the atmosphere. The apparent magnitude is calculated by removing this influence through observations of standard stars with well-defined apparent magnitudes. However, this is difficult if atmospheric conditions change on short timescales (shorter than the time required to undertake the observations). Alternatively, a good estimate of the apparent magnitude of a target can be calculated by comparing the instrumental magnitudes of other stars, in the field of view (FoV), to their known apparent magnitudes, and compensating for the difference.

We developed a software application that facilitates this process by matching observed stars in the FoV of the target to a catalogue and calculating the relationship between the two sets of magnitudes using a weighted least-squares approach. An existing algorithm, proposed by Groth[1] and Stetson[2], was used for the matching. The software decreases the statistical uncertainty in the calibrated magnitude relative to what can be achieved through a manual selection of a few stars by hand, by using a large number of stars.

The software was tested on infrared observations of a number of Be/X-ray binary systems in the Small Magellanic Cloud (SMC). A good agreement was found between our results and previous results for these systems. Additionally, using our software, we obtained smaller error margins in the calibrated magnitudes, compared to previous values calculated with the same data using similar techniques.

Be/X-ray binary systems

Be/X-ray binaries consist (in general) of a neutron star which is accreting material from the Be star it is orbiting (see Reig[3] for a review, and Sturm et al.[4] and Li et al.[5] for a possible exception). The Be star is surrounded by a circumstellar disc, as is evidenced by the presence of emission lines in the spectrum of these stars.[6] In addition to these emission lines, Be stars show a larger flux in the infrared than would be expected from a blackbody (stellar) emitter. This infrared excess is also believed to originate from the circumstellar disc and the Be star's spectrum can be fitted by including a free-free component in addition to the blackbody spectrum of the underlying star.[7-9] The size of the circumstellar disc is also variable as is reflected in the emission lines and in the infrared excess. For example, Telting et al.[10] showed the long term variability in X Persei between a disc and discless phase. Interferometry is now allowing for more detailed studies of these systems. See Rivinius et al.[11] for a recent review.

The X-ray emission observed from Be/X-ray binaries is the result of matter accreting onto the surface of the neutron star. Two types of outbursts are observed in these systems: Type I outbursts result when the neutron star moves through the circumstellar disc of the Be star during its orbit, while Type II outbursts are much brighter and longer lasting outbursts that arise when the circumstellar disc of the Be star is larger than normal and continuous accretion occurs. Because the Type II outbursts occur during periods of increased circumstellar disc size, this corresponds to periods of increased infrared flux from the system. Observations after a Type II flare show a decrease in the infrared flux.[12,13] Long-term studies of optical and infrared variability of Be/X-ray binaries (e.g. Rajoelimanana et al.[14]) are therefore an important tool in studying these systems.

In this paper, we used data obtained as part of long-term monitoring of Be/X-ray binaries with the IRSF telescope, to test the robustness of the software application developed. We report on our results in the sections that follow.

Data acquisition, reduction and analysis

The Infrared Survey Facility telescope

Infrared observations were undertaken with the 1.4-m Infrared Survey Facility (IRSF) telescope at the South African Astronomical Observatory (SAAO), located near Sutherland, Northern Cape, using the SIRIUS (Simultaneous 3 colour Infrared Imager for Unbiased Survey) camera. This camera allows simultaneous exposures to be taken at J (1.25 μm), H (1.65 μm) and K_s (2.15 μm) bands with a field of view of $\sim$7.7x7.7 arcmin^2.[15] A typical observation consisted of 15 dithers, each with an exposure time of 20 s.

Data for six Be/X-ray binaries in the SMC, gathered with the IRSF telescope during three observation campaigns that took place from 12 to 18 December 2007, from 30 December 2009 to 5 January 2010, and from 17 to 21 December 2010, were analysed and the results are presented in this paper.

Data analysis

The combination of the dithers and the photometric data reduction was done with the SIRIUS09[16] pipeline package that is integrated into the Image Reduction and Analysis Facility (IRAF) software package.[17]

Point Spread Function (PSF) photometry was performed on the science frames using the standard IRAF/NOAO packages.[18] The resulting photometric output of instrumental magnitudes was converted into a list of stars recording the *x* and *y* frame coordinates of the star, and the measured instrumental magnitudes with the associated (statistical) uncertainties. This list was used with the AutoCal program and the published IRSF catalogue[19] to calibrate the target's magnitude.

The AutoCal software application

A typical calibration workflow using AutoCal can be loosely divided into the three sections of preparation: calibration, post calibration analysis and fit correction. The steps undertaken during these three phases are briefly discussed below.

The preparation phase of AutoCal

In addition to the list of magnitudes determined by IRAF, the software takes as input the FITS (Flexible Image Transport System) file that was analysed. This allows the FITS image to be displayed and offers the user visual feedback of the matching process. Information is also read from the FITS header, including the Right Ascension (RA) and Declination (Dec) at the centre of the image and its width and height in pixels. The user is prompted for any additional information not included in the header that is required, such as the telescope's plate scale. The target star is selected by clicking on the displayed image or specifying its *x,y* coordinates.

In order to optimise the search for catalogue magnitudes and to ensure consistent formatting, catalogue data is imported into a database that is linked into AutoCal. This import is only performed the first time a catalogue is used. Once the catalogue has been added to the database, it will remain available during future uses of the program. All database tables contain the fields RA, Dec and magnitude, followed by a corresponding error-in-magnitude field. The interface prohibits the addition or omission of fields, ensuring that the required format is adhered to.

After the instrumental magnitudes are loaded into AutoCal, a database query is initiated to locate all stars in the catalogue that fall within the FoV of the image. The FoV is determined by the RA and Dec position from the FITS header, together with the image size (px) and plate scale (arcsec/px). The user may, alternatively, enter the ranges manually or specify only a percentage of the FoV to be used.

Calibration procedure

The calibration procedure comprises the following two main steps. Firstly, stars from the frame are matched to the corresponding catalogue stars, and secondly, a linear equation describing the relationship between the magnitudes of the pairs of matched stars is calculated and applied to the target star's instrumental magnitude. These two steps are described in the sections that follow.

Star matching

The pattern-matching algorithm implemented in the program was concurrently developed by Groth[1] and Stetson[2], with the former publishing it and the latter describing his version in lecture notes. A more detailed description of the algorithm can be found in the discussions by these authors. The algorithm starts by creating two lists of triangles, one containing all possible triangles (with stars as vertices) constructed from the stars in the frame star list and another containing all possible triangles constructed from those in the catalogue star list, and then searching for similarly shaped triangles from these two lists. The rationale behind the use of triangles is that the shape of a triangle is independent of translation, rotation, magnification and flips, and that corresponding star fields observed with different instruments and at different angles can therefore be matched easily. The algorithm then proceeds to pair vertices of matching triangles, or, in other words, stars.

In AutoCal, an initial matching is done on a user defined number of the brightest stars on the frame (default 25). From this initial matching, a coordinate transformation is calculated and applied to convert the coordinate system of the catalogue to that of the frame.

With the two sets of stellar coordinates now in the same coordinate system, entries in the two lists that are in close proximity are matched. At this stage, most of the stars in the frame list should be matched with a catalogue entry. This larger set of matches is then used to calculate and apply a new coordinate transformation between the two lists, effectively serving as a refinement to the initial transformation. Finally, all stars with positions that are in close proximity to the final transformed catalogue positions are matched to those entries. These pairs are used to perform the magnitude calibration, which is discussed in the next section.

Calculation of the calibrated magnitude of the target

Following the star matching, the best linear relationship between the magnitudes of the frame stars and their matches from the catalogue has to be calculated. We implemented a weighted least-squares approach to determine this relationship, which is applied to the instrumental magnitude of the target in order to obtain its calibrated apparent magnitude. This section motivates the rationale behind using a linear relationship and explains the weighted least-squares approach in more detail.

To a first-order approximation (see Bailer-Jones et al.[20] for an explanation of why a first-order approximation is used) for the extinction correction, the catalogue and instrumental magnitudes are related by:

$$m_{cat} = m_{ins} - k_\lambda X + \phi, \qquad \text{Equation 1}$$

where k_λ is the extinction correction coefficient (at wavelength λ), X is the airmass and ϕ is a zero-point offset between m_{cat} and m_{ins}. Therefore, as all observations on the same FoV occur at the same airmass (within less than 0.3%), we have calibrated the instrumental magnitudes against the catalogue magnitudes by assuming that the relation:

$$m_{cat} = g m_{ins} + c \qquad \text{Equation 2}$$

may be used, where g and c are two constants that can be determined by a least squares fit. The value of the gradient should, ideally, be $g = 1$ in all cases. However, we have included it as a free parameter and have used a weighted least squares (WLS) method to determine the values of g and c. Because the crowded star fields, obtained from the SMC observations generally contained a large portion of faint stars (with small signal to noise ratios) and only a small portion of bright stars (with large signal to noise ratios), this causes the fainter stars to skew the straight line fit if weights are not applied. The weight applied by AutoCal is calculated as:

$$w = \frac{1}{E^2_{m_{cat}} + E^2_{m_{ins}}} \qquad \text{Equation 3}$$

where $E_{m_{cat}}$ and $E_{m_{ins}}$ are the reported statistical errors in the magnitudes of the frame stars and catalogue stars (m_{cat} and m_{ins}) respectively. The matrix form of the weighted least squares system that must be solved is:

$$\begin{bmatrix} c \\ g \end{bmatrix} = \begin{bmatrix} \sum_i w_i & \sum_i (m_{ins_i} w_i) \\ \sum_i (m_{ins_i} w_i) & \sum_i (m^2_{ins_i} w_i) \end{bmatrix}^{-1} \begin{bmatrix} \sum_i (m_{cat_i} w_i) \\ \sum_i (m_{cat_i} m_{ins_i} w_i) \end{bmatrix} \qquad \text{Equation 4}$$

The 95% confidence interval for the apparent magnitude of the target was calculated using a two-sided Student's t value with $n - 2$ degrees of freedom, where n denotes the number of matched star pairs. It can be shown that the apparent magnitude lies within:

$$m^{target}_{app} \pm S\, t(0.975, n-2) \text{ with} \qquad \text{Equation 5}$$

$$S = X^T (Z^T W Z)^{-1} X \times \widehat{MSE} \qquad \text{Equation 6}$$

Here,

$$X = \begin{bmatrix} 1 \\ m_{ins,target} \end{bmatrix},\quad Z = \begin{bmatrix} 1 & m_{ins1} \\ 1 & m_{ins2} \\ \vdots & \vdots \\ 1 & m_{insn} \end{bmatrix},\quad W = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix},\quad \text{and}$$

$$\widehat{MSE} = \sum_{i=1}^{n} (w_i^{1/2} m_{cat_i} - w_i^{1/2} \widehat{m_{cat_i}}),$$

where $\widehat{m_{cat_i}} = g m_{ins_i} + c$.

Analysis of calibration results

Following completion of the calibration, the magnitude and uncertainty in the magnitude is determined for the target. The program displays a plot of m_{cat} vs. m_{ins} with the data points, the fitted line and other information relevant to the fit. The user is able to revise this calibration graph by choosing whether points should be excluded from the WLS fit. This is an interactive process and the user is able to select and/or remove stars interactively from the plot (a change in colour indicating which stars are included) by either clicking on individual stars, or by dragging a region around the stars that should be effected. This automatically updates the WLS fit, displaying the new best fit line on the data the equation of the fit, and re-calculates the target's apparent magnitude. The interactive nature of the star exclusion/inclusion allows for a rapid first calibration of the data.

Results

In total, AutoCal was tested on 15 observations of Be/X-ray binaries, each including J, H and K_s bands, and successfully calculated apparent magnitudes for all 43 FITS frames for which the target could be detected (for two K_s frames the target star was too faint to detect). The procedure we followed to obtain the apparent magnitudes was similar for all the data frames, so only one frame will be discussed in detail. The results of the calibrations performed by AutoCal are then summarised and finally, our results for a single observation are compared with previously published results from the same observation.

The J-band observation of SXP2.37 on 14 December 2007

SXP2.37 was observed on 14 December 2007 in J, H and K_s, with 15 dithers, each with an exposure time of 20 s. Here follows a description of how we performed the magnitude calibration for the J-band observation.

Star matching

Photometry with IRAF identified and measured the magnitudes of 587 stars in the field of view around the target. The database query for the region found 1417 stars, of which 1307 remained after stars closer to one another than the elimination threshold were removed. The initial position match was attempted with the brightest 25 stars in both lists and the matching algorithm found 15 matched pairs that were used to calculate the initial transformation. Subsequent analysis with all 587 star positions matched a total of 493 pairs.

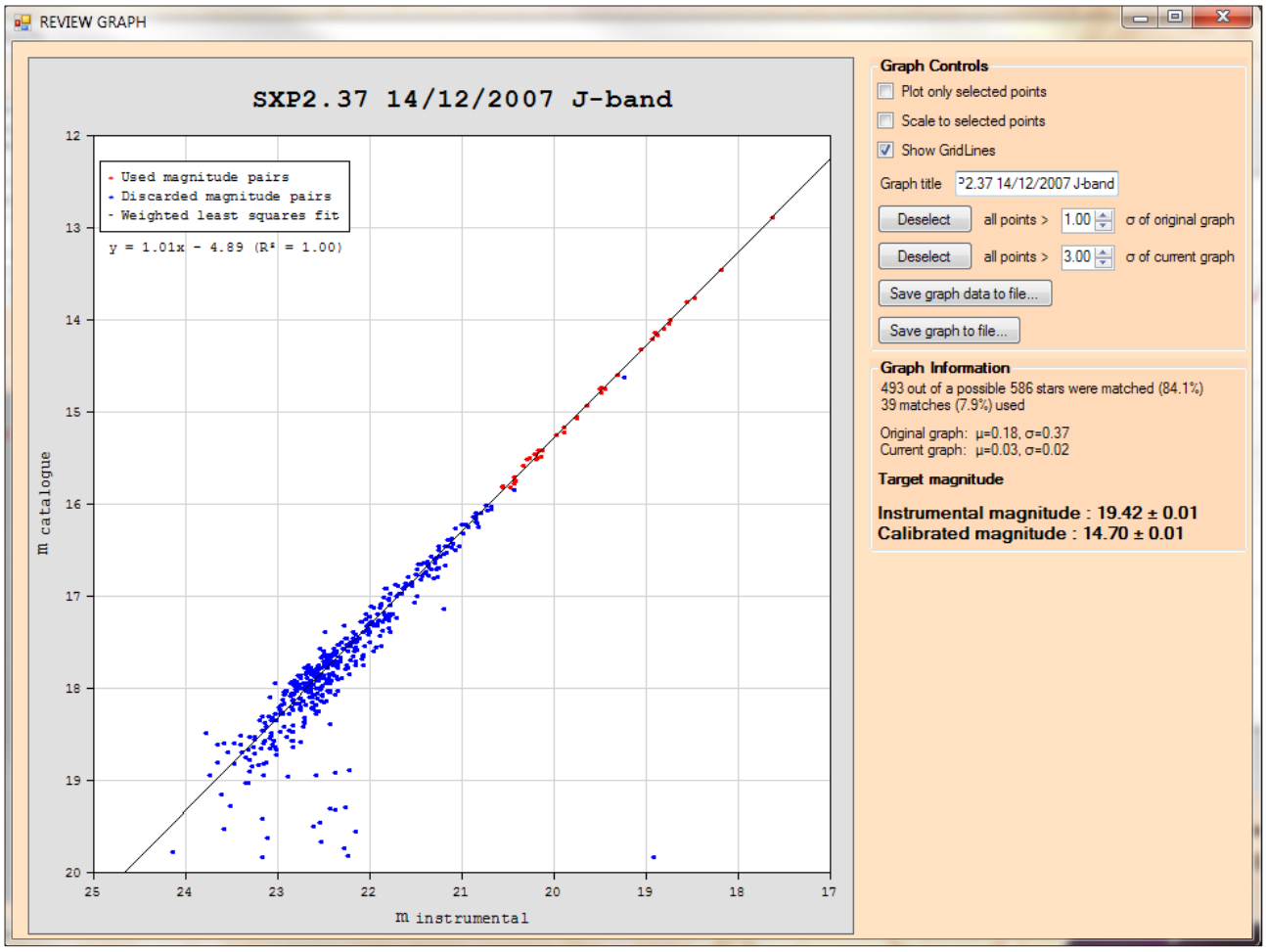


Figure 1: Calibration graph for the J-band observation of SXP2.37 on 14 December 2007.

Magnitude calibration

Figure 1 shows the calibration graph for the J-band observation of SXP2.37 after it was revised by the user. The solid line represents the WLS solution through the data points. The apparent magnitude calculated by AutoCal was 14.70±0.02 mag with initial values r^2=0.85 and a slope of g=1.02 (all uncertainties in the magnitude are recorded at the 95% confidence level).

The scatter in the data is clearly larger for fainter stars. Because of this, and because some stars may still be false matches, have incorrect instrumental or catalogue magnitudes measurements, or have presented intrinsic variability, we tested the robustness of the fit by restricting the fit to magnitude values near the range of the target. We excluded data points from the calibration graph by first removing all faint stars with m_{ins} greater than a user selected cut-off value, then removing any obvious outliers and finally removing all points more than 3σ from the straight line fit.

For the example discussed here, all stars fainter than $m_{ins}\geq$20.9 mag were removed, after which the one obvious outlier (m_{ins}=18.9 mag, m_{cat}=19.9 mag) and two stars that were more than 3σ from the straight line fit were discarded. This removed 441 of the original 493 matches (the blue dots in Figure 1). The selection of the 20.9 mag cut-off point was based on a visual inspection and we found that for this frame, applying a cut-off of any value between 20 mag and 22 mag yielded very similar results. The apparent magnitude that was calculated using the smaller sample was 14.70±0.01 mag with r^2=1.00 and a slope of g=1.01. For this J-band observation of SXP2.37 there was no significant change in the calibrated apparent magnitude when using the smaller sample of matches, demonstrating that the WLS method is robust despite outliers. However, for some of the other frames analysed with this method, we found larger discrepancies between the WLS fits undertaken before and after the removal of outlier data points, especially for certain frames that contained a large number of outliers. For those frames, removing the obvious outliers and faint stars with scattered magnitudes and lower signal to noise ratios, resulted in better linear fits near the magnitudes of the target stars and noticeable changes in the calibrated magnitudes.

Results of calibration for 43 frames

Table 1 lists the calibration results for the six sources analysed in this study. For every frame but one, matches were found for at least 70% of the available stars from the frame star list. There was one exception where only 45 stars were available for matching and only slightly more than half of the stars could be matched by the algorithm.

Table 1: Auto-calibration results for six Be/X-ray binaries in the Small Magellanic Cloud.

Target	Date or campaign	Stars on frame			Number of matched stars			Matches used			Gradient (g)			Apparent magnitudes		
		J	H	K_s	J	H	K_s	J	H	K_s	J	H	K_s	J	H	K_s
SXP0.72	2MASS	–	–	–	–	–	–	–	–	–	–	–	–	13.45±0.02	13.51±0.03	13.47±0.05
	18/12/2007	92	86	54	78 (85%)	67 (78%)	42 (78%)	28	24	14	1.02	1.03	1.01	13.58±0.03	13.62±0.02	13.64±0.02
	05/01/2010	93	86	45	65 (70%)	62 (72%)	23 (51%)	27	27	15	1.01	1.01	1	13.56±0.03	13.60±0.02	13.61±0.02
SXP2.37	2MASS	–	–	–	–	–	–	–	–	–	–	–	–	14.77±0.05	14.90±0.09	14.69±0.13
	14/12/2007	586	471	213	493 (84%)	404 (86%)	177 (83%)	57	54	34	1.01	1.02	1.01	14.70±0.01	14.68±0.01	14.63±0.01
	05/01/2010	358	376	123	327 (91%)	340 (90%)	111 (90%)	59	62	36	1	1.01	1.01	14.99±0.01	14.96±0.01	14.94±0.02
SXP2.76	2MASS	–	–	–	–	–	–	–	–	–	–	–	–	14.04±0.04	13.91±0.04	13.66 (L)
	17/12/2007	293	248	182	246 (84%)	215 (87%)	145 (80%)	44	60	25	1.01	1	1	14.07±0.02	14.00±0.01	13.80±0.01
	05/01/2010	262	160	97	237 (91%)	147 (92%)	80 (83%)	49	45	26	1.01	1	0.99	13.93±0.02	13.83±0.01	13.68±0.02
	21/12/2010	44	100	46	42 (96%)	96 (96%)	44 (96%)	27	34	14	0.99	1.01	0.99	13.90±0.04	13.79±0.01	13.62±0.04
SXP7.92	2MASS	–	–	–	–	–	–	–	–	–	–	–	–	13.84±0.03	13.81±0.04	13.66±0.04
	30/12/2009	501	425	215	453 (90%)	387 (91%)	183 (85%)	52	51	30	1.01	1.01	1.01	13.72±0.01	13.63±0.01	13.51±0.01
	17/12/2010	306	230	102	286 (94%)	225 (98%)	93 (91%)	34	50	22	1	1	1	14.30±0.01	14.16±0.01	13.96±0.02
SXP11.5	2MASS	–	–	–	–	–	–	–	–	–	–	–	–	14.16±0.03	13.70±0.04	13.52±0.04
	30/12/2009	80	61	41	73 (91%)	53 (87%)	35 (85%)	67	51	33	0.99	1	1.01	14.25±0.01	13.75±0.01	13.61±0.01
	19/12/2010	248	229	110	221 (89%)	211 (92%)	90 (82%)	39	49	22	1	1.01	1.02	14.37±0.01	13.85±0.01	13.69±0.02
SXP18.3	2MASS	–	–	–	–	–	–	–	–	–	–	–	–	15.87±0.10	15.12 (L)	15.58±0.24
	30/12/2009	804	907	372	744 (93%)	832 (92%)	342 (92%)	94	87	66	1	1.01	1.02	14.85±0.00	14.74±0.01	14.68±0.01
	19/12/2010 (A)	493	421	–	479 (97%)	409 (97%)	–	62	74	–	1	1	–	16.11±0.02	15.96±0.02	–
	19/12/2010 (B)	642	433	155	612 (95%)	421 (97%)	146 (94%)	65	69	48	1	1	1.01	16.10±0.02	16.00±0.02	16.24±0.05
	19/12/2010 (C)	511	421	–	494 (97%)	405 (96%)	–	62	69	–	1	1	–	16.11±0.02	16.19±0.02	–

(L) Indicates that an accurate magnitude could not be calculated during the 2MASS campaign and the corresponding value is a lower bound for the magnitude.

The largest uncertainty in the calculated apparent magnitude values was 0.05 mag. For all 43 WLS fits $r^2 \approx 1$, and the gradient lay between $0.99 \leq g \leq 1.03$. This lends strong support to the simplifying assumption made in equation 2.

Comparison with previously published results for SXP11.5

Townsend[13] and Townsend et al.[21,22] published calibrated magnitudes for the same observations of SXP11.5 on 30 December 2009 that we present here. They also used a large number of stars in the FoV to perform the calibration. Their calibrated magnitudes were J=14.23±0.04 mag, H=13.75±0.03 mag, K_s=13.61±0.04 mag, while we calculated J=14.25±0.01 mag, H=13.75±0.01 mag, K_s=13.61±0.01 mag. It is clear that the two sets of results all agree within the specified uncertainties, but that the uncertainties in our results are smaller throughout.

A possible explanation for the differences became apparent when we repeated our calibration for the J-band without performing outlier removal, the results of which can be seen in the left half of Figure 2. The new result was a J-band magnitude of 14.24±0.04 mag. The uncertainty now matches that of the previously published results exactly and the magnitude is also closer to theirs. A plausible explanation for our uncertainties being smaller is our use of a weighted least-squares fitting that assigns smaller weights to the majority of fainter stars on the frames, coupled with the interactive outlier removal based on the appearance of the WLS plot.

Discussion

A number of the sources considered here have been previously discussed by other authors. For example, Rajoelimanana et al.[14] presented the long-term I-band variability of 38 Be/X-ray binaries in the SMC, some of which include the time period covered by the observations presented here. In addition, some of the IRSF data used in this study have been previously analysed and presented by Townsend[13] and Townsend et al.[213,22], as was discussed earlier in the paper. Furthermore, the magnitudes calculated in this study are compared with those from the 2 Micron All Sky Survey (2MASS).[23] The times at which the observations were performed are reported in Modified Julian Date (MJD). Our results for each source are discussed individually below, and brought into the context of the existing literature.

SXP0.72

SXP0.72 was observed on 18 December 2007 (MJD 54 452.77) and 5 January 2010 (MJD 55 201.79). As can be seen from Figure 3, a comparison of the magnitudes calculated from the two observations showed no significant variability in any of the bands. Furthermore, only a very slight change was found between the 2MASS magnitudes and those calculated here (see Table 1), corresponding to a decrease in flux of between 10% and 15% for each of the three bands.

SXP2.37

SXP2.37 was observed on the 14 December 2007 (MJD 54 448.83) and 5 January 2010 (MJD 55 201.81). The calibrated magnitudes are comparable to previous 2MASS magnitudes observed for the system (J=14.77±0.05 mag, H=14.90±0.09 mag, K_s=14.69±0.13 mag). There is a significant, albeit small, increase in magnitude of ~0.3 mag observed between the first and second observing runs. This slight variation is indicative of the variability associated with the circumstellar disc. This variability is clearly shown in Figure 4a of Rajoelimanana et al.[14], that demonstrates the long-term variation.

SXP2.76

SXP2.76 is another source that showed very little infrared variation during all three observing runs, and the magnitudes are comparable to the 2MASS observations (see Table 1 and Figure 3). A very slight brightening was observed in all bands between the first and second runs (MJD 54 451.78 and 55 201.84) corresponding to an increase in flux of between 11% and 17%. The magnitudes during the third run (MJD 55 551.88) were similar to those for the second run. Rajoelimanana et al.[14] showed that SXP 2.76 shows magnitude variation, in I band, with a period of roughly 2800 days and an amplitude of ~0.2 mag (see their Figure 5a). The increase in brightness we observed in the infrared is consistent with these results.

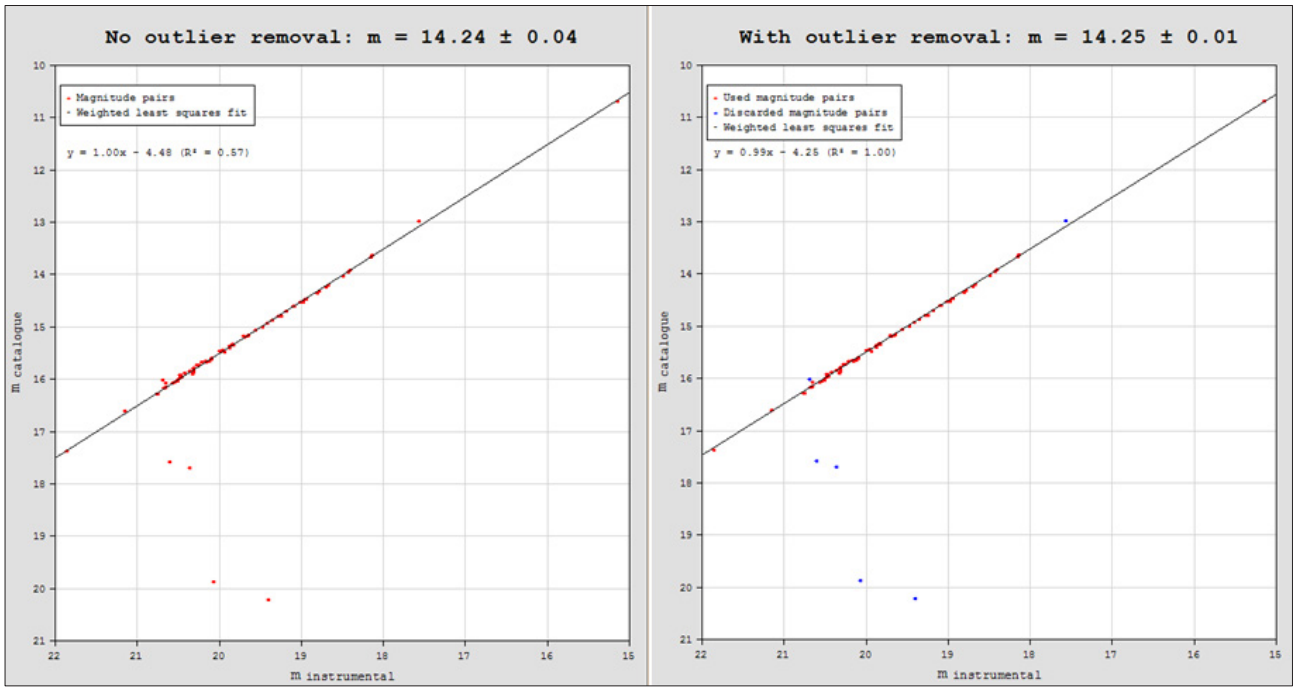


Figure 2: Calibration graphs for SXP11.5 showing the calibrated magnitudes, before (left) and after (right) outlier removal.

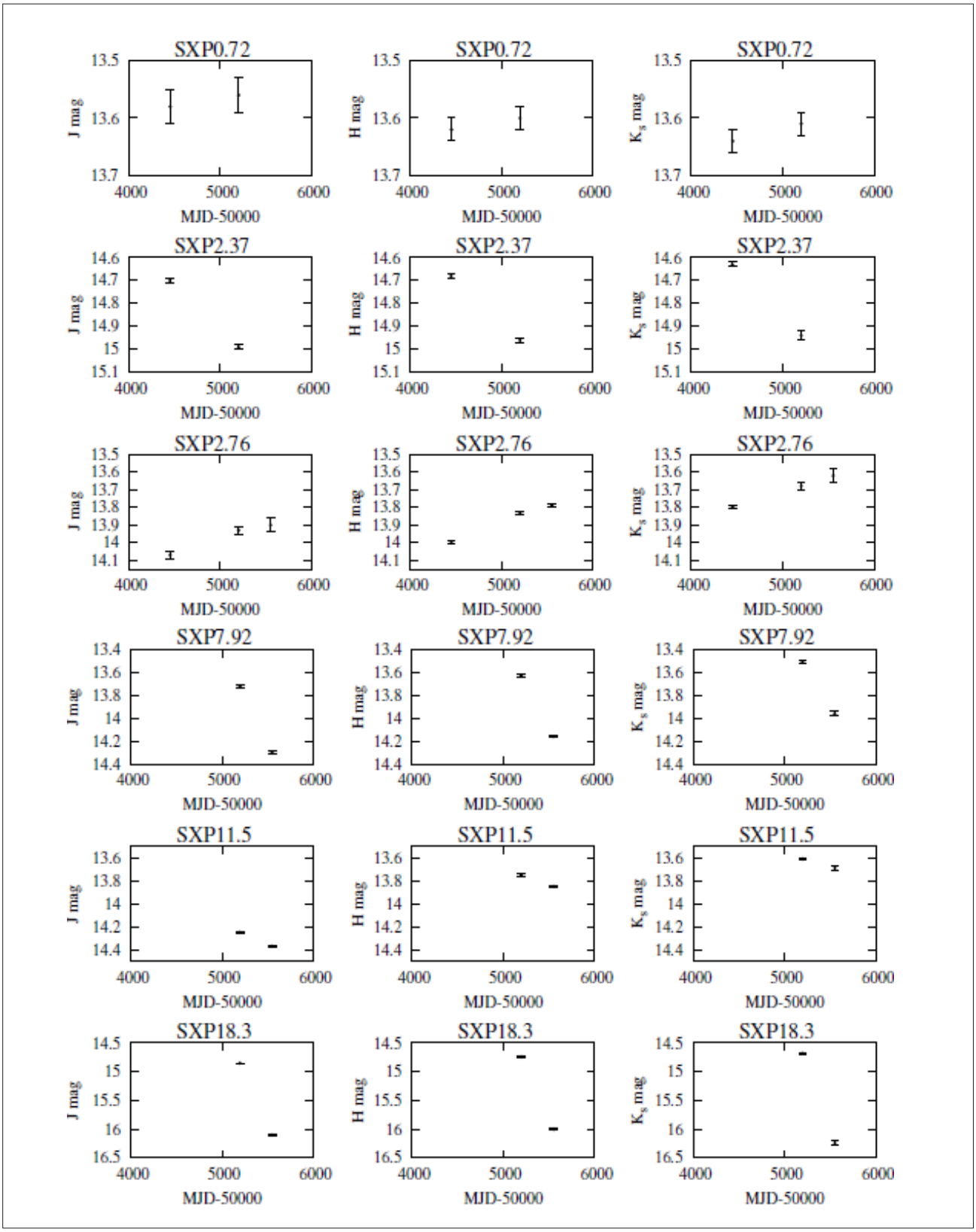


Figure 3: Lightcurves for six Be/X-ray binaries in the Small Magellanic Cloud.

SXP7.92

SXP7.92 was observed during the second and third observing runs, on MJD 55 195.81 and MJD 55 547.78, respectively. The system was very slightly brighter during the first run when compared to the 2MASS catalogue, but is much fainter during the third run in all three bands (see Table 1 and Figure 3).

This clearly indicates the variable nature of this system, and the decrease in infrared flux during the third observation run indicates a decrease in

the size of the circumstellar disk, consistent with the variability reported in Rajoelimanana et al.[14]

SXP11.5

SXP11.5 was observed on MJD 55 195.82 and MJD 55 549.76. A small, though significant variability (~0.1 mag) was detected and the magnitudes are comparable to the 2MASS catalogue.

As shown previously, our results are in good agreement with those that Townsend[13] calculated from the same observation on 30 December 2009.

SXP18.3

Of the six sources covered by this study, SXP18.3 showed by far the highest level of variability in the near infrared. It was observed on MJD 55 195.85 and 3 times consecutively around MJD 55 549.80. We used the magnitude calculated from the second of these observations (listed as observation B in Table 1), because it was the only one for which the target could be identified in the K_s-band. A change in magnitude of 1.02 mag and 0.9 mag was observed from the time of the 2MASS campaign to the second observing run for the J and K_s-band respectively. The J and K_s-band brightening is equivalent to flux increases of 156% and 129% respectively. An even greater variability was found between the second and third observation runs, equating to a decrease in flux of 68%, 69% and 76% in J, H and K_s respectively. The magnitudes calculated from the data captured during the second and third observation runs are indicated in Figure 3.

From Figure 14a in Rajoelimanana et al.[14], the optical data clearly show an extended period of brightening, starting around MJD 52 800 and lasting until around MJD 54 600, a period of about five years. The drastic decrease in the brightness of the system from December 2009 to December 2010 seems consistent with the aftermath of another Type II outburst. This idea is supported by the same figure, that shows a sharp I-band brightening for the period around MJD 54 700 to MJD 55 000, which leads up to the IRSF observation that was taken on MJD 55 549.

Conclusion

Of the 45 FITS frames that were created from the data captured during the three IRSF observations, the target was visible on 43. AutoCal was successful in calculating calibrated magnitudes for all of these frames. The correctness of these magnitudes is demonstrated by the good agreement with previously published results for the same data. That section further argued that the interactive outlier removal through inspection of the WLS plot, together with the use of WLS methods to calculate uncertainties, allows for smaller uncertainties in the calibrated magnitudes.

AutoCal proved robust in the sense that it managed to successfully calibrate all the frames that we tested, irrespective of their stellar density. Despite the fact that all the frames can be said to be reasonably densely populated, the number of stars they contained ranged from 41 to 907. All the calibrations were performed with the user-specified parameters at their default values. The fact that AutoCal's functioning was not sensitive to these values ensured faster calibration and a good user experience. AutoCal has the potential to be expanded further to allow blind searches for variability to be undertaken.

We were able to detect statistically significant infrared variability for each of the six Be/X-ray binaries discussed here. The results for SXP11.5 agree well with those reported in Townsend[13], while those of SXP2.37 and SXP18.3 point towards the same events that caused the variability seen in the OGLE and MACHO data presented in Rajoelimanana et al.[14]

Acknowledgements

The authors are grateful to M.J. von Maltitz (UFS) for all his help regarding the statistics of the error analysis and for assistance, advice and comments from P.J. Meintjes (UFS) and P.J. Blignaut (UFS). The IRSF data used in this study were in part collected by L.J. Townsend, and the authors thank him for making it available. This paper uses observations made at the South African Astronomical Observatory. DJW undertook this research in part under funding provided by the National Astrophysics and Space Science Programme. This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. The authors would like to thank the reviewers of this paper for meaningful suggestions that contributed to the quality of the final version.

Authors' contributions

DJW wrote the software application, performed the data reduction and wrote the majority of the manuscript. BvS captured some of the data with the IRSF telescope and ensured the correctness of the more technical parts of the manuscript, especially the section on Be/X-ray binaries. BvS also wrote some sections of the manuscript.

References

1. Groth EJ. A pattern-matching algorithm for two-dimensional coordinate lists. Astron J. 1986;91(5):1244–1248. http://dx.doi.org/10.1086/114099
2. Stetson PB. The techniques of least squares and stellar photometry with CCDs [article on the Internet]. c2000 [updated 2000 Jun 21; cited 2011 Jun 26]. Available from: http://ned.ipac.caltech.edu/level5/Stetson/Stetson1_3_3.html
3. Reig P. Be/X-ray binaries. Astrophys Space Sci. 2011;332(1):1–29. http://dx.doi.org/10.1007/s10509-010-0575-8
4. Sturm R, Haberl F, Pietsch W, Coe MJ, Mereghetti S, La Palombara N, et al. A new super-soft X-ray source in the Small Magellanic Cloud: Discovery of the first Be/white dwarf system in the SMC? Astron Astrophys. 2012;537:A76. http://dx.doi.org/10.1051/0004-6361/201117789
5. Li KL, Kong AKH, Charles PA, Lu TN, Bartlett ES, Coe MJ, et al. A luminous Be+ white dwarf supersoft source in the wing of the SMC: MAXI J0158-744. Astrophys J. 2012;761(2):99–111. http://dx.doi.org/10.1088/0004-637X/761/2/99
6. Struve O. On the origin of bright lines in spectra of stars of class B. Astrophys J. 1931;73:94–103. http://dx.doi.org/10.1086/143298
7. Woolf NJ, Stein WA, Strittmatter PA. Infrared emission from Be stars. Astron Astrophys. 1970;9:252–258.
8. Waters LBFM, Lamers HJGLM. The infrared and radio spectrum of early type stars with mass loss. II. Tables of theoretical curves of growth for IR and radio excess and gaunt factors. Astron Astrophys. 1984;57:327–52.
9. Hummel W, Vrancken M. Line formation in Be star circumstellar disks. Shear broadening, shell absorption, stellar obscuration and rotational parameter. Astron Astrophys. 2000;359:1075–1084.
10. Telting JH, Waters LBFM, Roche P, Boogert ACA, Clark JS, De Martino D, et al. The equatorial disc of the Be star X Persei. Mon Not R Astron Soc. 1998;296(4):785–799. http://dx.doi.org/10.1046/j.1365-8711.1998.01433.x
11. Rivinius T, Carciofi AC, Martayan C. Classical Be stars. Rapidly rotating B stars with viscous Keplerian decretion disks. Astron Astrophys. 2013;21(1).
12. Smith MA, Henrichs HF, Fabregat J. Be stars in X-ray binary systems. In: Coe MJ, editor. The Be phenomenon in early-type stars; IAU Colloquium 175, 1999 Jun 28–Jul 2; Alicante, Spain. San Francisco, CA: ASP Conference Series; 2000.
13. Townsend LJ. An X-ray, optical and infra-red study of high-mass X-ray binaries in the Small Magellanic Cloud [PhD thesis]. Southampton: University of Southampton; 2012.
14. Rajoelimanana AF, Charles PA, Udalski A. Very long-term optical variability of high-mass binaries in the Small Magellanic Cloud. Mon Not R Astron Soc. 2011;413(3):1600–22. http://dx.doi.org/10.1111/j.1365-2966.2011.18243.x
15. Nagata T, Tanabé T, Glass IS. IRSF observing manual [user manual on the Internet]. c2003 [cited 2013 Jan 7]. Available from: http://www.kusastro.kyoto-u.ac.jp/~nagata/Irsf/IRSFman_e15.pdf.
16. Nakajima Y. SIRIUS09 pipeline users manual [manual on the Internet]. c2011. [cited 2013 Jan 7]. Available from: http://irsf-software.appspot.com/yas/nakajima/sirius/sirius09.html

17. Image Reduction and Analysis Facility (IRAF). IRAF Version 2.16 [computer program on the Internet]. c2015 [cited 2013 Jan 8]. Available from: http://iraf.noao.edu/

18. Massey P, Davis LE. A user's guide to stellar CCD photometry with IRAF [manual on the Internet]. c1992 [cited 2013 Jan 8]. Available from: https://www.mn.uio.no/astro/english/services/it/help/visualization/iraf/daophot2.pdf

19. Kato D, Nagashima C, Nagayama T, Kurita M, Koerwer JF, Kawai T, et al. The IRSF Magellanic Clouds point source catalog. Publ Astron Soc Jpn. 2007;59(3):615–641. http://dx.doi.org/10.1093/pasj/59.3.615

20. Bailer-Jones CA, Lamm M. Limits on the infrared photometric monitoring of brown dwarfs. Mon Not R Astron Soc. 2003;339(2):477–485. http://dx.doi.org/10.1046/j.1365-8711.2003.06189.x

21. Townsend LJ, Coe MJ, McBride VA, Bird AJ, Schurch MPE, Corbet RHD, et al. Be/X-ray binary SXP6.85 undergoes large Type II outburst in the Small Magellanic Cloud. Mon Not R Astron Soc. 2010;403(3):1239–1245. http://dx.doi.org/10.1111/j.1365-2966.2009.16211.x

22. Townsend LJ, Coe MJ, Corbet RHD, McBride VA, Hill AB, Bird AJ, et al. The orbital solution and spectral classification of the high-mass X-ray binary IGR J01054-7253 in the Small Magellanic Cloud. Mon Not R Astron Soc. 2011;410(3):1813–1824. http://dx.doi.org/10.1111/j.1365-2966.2010.17563.x

23. Skrutskie MF, Cutri RM, Stiening R, Weinberg MD, Schneider S, Carpenter JM, et al. The Two Micron All Sky Survey (2MASS). Astron J. 2006;131(2):1163–1183. http://dx.doi.org/10.1086/498708

Fertigation of *Brassica rapa* L. using treated landfill leachate as a nutrient recycling option

AUTHORS:
Frank O. Alaribe[1]
Pariatamby Agamuthu[1]

AFFILIATION:
[1]Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia

CORRESPONDENCE TO:
Frank Alaribe

EMAIL:
alaribe@gmail.com

POSTAL ADDRESS:
Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia

DATES:
Received: 05 Feb. 2015
Revised: 19 July 2015
Accepted: 25 Aug. 2015

KEYWORDS:
fertigation; landfill; treated leachate; *Brassica rapa* L; nutrient

HOW TO CITE:
Alaribe FO, Agamuthu P. Fertigation of *Brassica rapa* L. using treated landfill leachate as a nutrient recycling option. S Afr J Sci. 2016;112(3/4), Art. #2015-0051, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20150051

Optimising nutrient availability and minimising plant metal contamination are vital in sustainable agriculture. This paper reports experiments in which treated leachate was used at different concentrations with predetermined N content for fertigation of *Brassica rapa* L. (leafy vegetable). An inorganic fertiliser, with N content equivalent to the leachate amount, was also prepared, as well as a control. Growth (leaf length, leaf width and stem height), harvest parameters (total number of leaves, root length and root dry weight) and specific growth rates (mm/day) were determined for three consecutive seasons. The dry weights of leaves, roots and stems in the leachate treatments were within the ranges of 1.95–3.60 g, 1.18–3.60 g and 0.33–1.37 g, with biomasses of 1.75 g, 1.14 g and 0.2 g, respectively, which were higher than those of the control. *B. rapa* L. fertigated with 25% diluted treated leachate recorded high specific growth rate and a leaf length of 0.53 mm/day and 0.23.17±0.58 cm, respectively (%N=0.023; $p<0.05$). The maximum permissible mineral concentration set by the Food and Agricultural Organization of the United Nations/World Health Organization (FAO/WHO) was compared with that of the grown plants. Treated leachate can increase plant nutrient content.

Introduction

Fertigation is the technique of supplying dissolved fertiliser to crops through irrigation systems. This approach can improve application efficiency and reduce the need to export fertiliser off-field.[1,2] Current nutrient management practices using chemical fertilisers exhibit adverse environmental effects. Continuous production of crops under intensified agriculture over years, has resulted in large-scale removal of nutrients from soil, thereby causing a negative nutrient balance and declining soil fertility.[3] Hence, an adequate proportion of nutrient supply, especially in the form of bio-fertilisers, to soil and/or crops is essential for enhancing global food security.[4] Long-term irrigation with wastewater may result in the accumulation of heavy metals in soils and plants, consequently decreasing food safety and increasing health risk, which are environmental concerns.[5,6] Health risks caused by heavy metal contamination of soil have been widely reported.[7–12] Crops and vegetables grown in soils contaminated with heavy metals may have a larger accumulation of metals than those grown in uncontaminated soils. Risk assessment of heavy metal accumulation in wastewater irrigated leafy vegetables, such as palak (*Beta vulgaris* L.), amaranthus (*Amaranthus caudatus* L.), and cabbage (*Brassica oleracea* L.) was reported by Anita et al.[6] Intake of metal-contaminated vegetables is a pathway for heavy metal toxicity to humans.[13,14] Intawongse and Dean[15] determined the bioavailability of Cd, Cu, Zn and Mn in the human gastrointestinal tract from the edible part of the vegetables by using an in-vitro gastrointestinal extraction technique. They found that the edible portions of 'lettuce and radish' are more responsible than other vegetables for heavy metal accumulation in humans.

Malaysia has approximately 260 landfills, of which 10% are engineered (safe collection and leachate treatment facilities) and 90% are un-engineered.[16] Agamuthu[16] indicated that the volume of leachate generated is about 3 million litres per day. Leachate is produced by precipitation and percolation of 'biodegradable and non-biodegradable' wastes deposited in landfills. Landfill leachate mainly consists of large amounts of organic matter, including dissolved organic matter, phenols, ammoniacal nitrogen, phosphates, heavy metals, sulfides, solids, inorganic salts, and other toxicants, and is characterised by hardness, acidity, alkalinity, and salinity.[16,17] The complex pollution and toxicity impacts of leachates have imposed the need to standardise pre-treatments prior to discharge for quality improvement.[18] Treated leachates contain nutrients and minerals that can possibly be transformed for fertigation. Scholars have reported N transport and distribution characteristics in soils for basin irrigation using fertigation and conventional fertilisation.[19-21,5] Cheng and Chu[22] reported the positive and detrimental effects of landfill leachate on plant growth, depending on plant species leachate concentration. Menser et al.[23] explained that irrigation with leachate can lead to yield reduction, leaf damage, premature senescence and poor plant survival. By contrast, Liang et al.[24] suggested that the use of landfill leachate as irrigation water in dry seasons can enhance the growth, survival, and stomatal conductance of *Acacia confusa*, *Leucaena leuocephala* and *Eucalyptus torelliana*. Cureton et al.[25] reported significantly high growth rates of *Phalaris arundinacea*, *Salix babylonica* and *Populus nigra* subjected to leachate application; however, phytotoxicity symptoms, such as brown leaves and necrotic spots, were observed in poplar leaves. Mapanda et al.[26] reported that more than 100 ha of land under horticultural production in Harare utilises wastewater for irrigation of crops such as maize (*Zea mays*) and leafy vegetables (*Brassica* spp.). Growth rate and biomass production are common indicators of imposed stress. Although leaf length, leaf width and shoot length exhibit fluctuating asymmetry, non-directional deviation from anticipated symmetry is proposed as an environmental stress indicator.[27,28] Nominal consumers consider undamaged, dark green and large leaves as characteristics of good-quality leafy vegetables; however, the external morphology of vegetables cannot guarantee safety from contamination. In particular, heavy metals rank the highest among the chief contaminants of leafy vegetables.[26,29,30]

Leafy vegetables grow in places with adequate water supply and well-drained, fertile and preferably alkaline soils. These plants prefer a pH range of 5.5–8.2 although they can tolerate a pH within the range of 4.3–7.5.[31] Leafy vegetables prefer cool, moist and reasonably fertile soils. Shallow rooted plants are intolerant of drought and thus need to be grown in moist, fertile soil to produce high-quality leaves.[32] *Brassica* and mustards need adequate nitrogen and sulfur[33] and an N:S ratio ranging from 4:1 to 8:1 is appropriate for the *Brassica* species. Treated landfill leachate can be utilised as a source of water or plant nutrients, and as a soil conditioner for crop production.

This study aimed to characterise leachate from the Ampar Tenang landfill and compare its performance with conventional inorganic fertiliser at different leachate concentrations for fertigation. The yields of *Brassica* at specific N concentrations in leachates and inorganic fertiliser were analysed to ascertain the effects of nutrients on plants and to compare them with the permissible concentration standards set by the Ministry of Agriculture, Fisheries and Forest, United Kingdom and the Food and Agricultural Organization of the United Nations/World Health Organization (FAO/WHO).

Materials and method

Site description

The Ampar Tenang landfill, located in the Sepang district of Selangor, Malaysia, began its operations in the year 2000 and was closed on 26 January 2010. Under the management of Alam Flora Sdn. Bhd., a waste management company, the 4-ha landfill received about 150 tonnes of domestic waste daily from Sepang and its environs. Leachate from this landfill indiscriminately leached into nearby un-engineered 'erosion' holes after precipitation.

Analysis of leachates

Specific characteristics of the leachate are displayed in Table 1. The pH and conductivity of the leachate were measured using a pH and conductivity probe (Hanna Model, No. 8033, Clarkson Laboratory, Chula Vista, CA, USA.). Total suspended solids and colour were determined using a spectrophotometer (HACH Model, DR/4000, Hach Company, Loveland, CO, USA). Biochemical Oxygen Demand (BOD_5), Chemical Oxygen demand (COD), and total-N were analysed according to the standard methods of the America Public Health Association (APHA), the America Works Association (AWWA) and the Water Environment Federation (WEF), respectively.[34] Heavy metals were determined in digested leachate by inductively coupled plasma optical emission spectrometry. The leachate was chemically treated using a Jar Test which involves a six-paddle flocculator, from Stuart Scientific (Stone, UK) equipped with 6 beakers of 500 mL each. Ferric (III) chloride at 4 g/L in solid state was used as effective coagulant dosage for raw leachate at pH 7.[35] The filtrate was used for fertigation treatments.

Table 1: Characteristics of untreated leachate samples from Ampar Tenang landfill compared with Environmental Quality (Control of Pollution from Solid Waste Transfer Station and Landfill) Regulation 2009 Malaysia[49]: Acceptable conditions for discharge of leachate.

Parameters	Units	(Ampar Tenang landfill leachate)	Standard	Method
Temperature	°C	29.8± 0.031	40	APHA2550B
pH at 25°C	–	6.58± 0.522	6.0-9	0APHA4500-H^+B
BOD_5	mg/L	209± 2.101	20	APHA2550B
COD	mg/L	3150± 3.521	400	APHA2550C
BOD_5/COD ratio	mg/L	0.07± 0.005	–	–
Chloride	mg/L	2671±18	NA	HACH8113
Turbidity	FAU	1260± 2.610	NA	APHA2540D
Sulfate	mg/L	221± 1.501	0.5	HACH 8131
Total N	mg/L	900± 0.00	5.0	APHA4500-N
Suspended Solid	mg/L	1718± 2.120	50	APHHA 2540D
Oil and Grease	mg/L	411.5± 0.001	5.0	APHA5520B
K	mg/L	3575±2.531	N.A	APHA3120B
Ca	mg/L	85±4.021	N.A	APHA3120B
Mg	mg/L	18.8±0.232	N.A	APHA3120B
Na	mg/L	1352±1.012	N.A	APHA3120B
Pb	mg/L	0.25±0.001	0.10	APHA3120B
Cd	mg/L	0.000001±0.001	0.01	APHA3120B
Se	mg/L	1.6±0.011	0.02	APHA3120B
Al	mg/L	19±1.201	N.A	APHA3120B
Mn	mg/L	6.75±0.170	0.2	APHA3120B
Cu	mg/L	0.245±0.011	0.2	APHA3120B
Zn	mg/L	2.35±0.003	2.0	APHA3120B
Fe	mg/L	15±0.111	5.0	APHA3120B
As	mg/L	0.3±0.002	0.05	APHA3120B
P	mg/L	45± 0.00	N.A	HACH8048

N.A, not available; BOD, Biochemical Oxygen Demand; COD, Chemical Oxygen Demand

Experimental design and growth conditions

The study was conducted in a designed greenhouse (1.2 m x 1.2 m x 3.3 m) near the IGS building, University of Malaya. *Brassica rapa* L. were grown in free-draining 4-L nursery poly bags, each containing 5 kg of soil, classified as Mollisol[36] (Table 2). Ten seeds were sown in germination pots to which distilled water of pH 6.94 (without metal contaminants) was applied daily. At day 6 after germination, uniform seedlings were selected and transplanted at a rate of two plants per bag (0.2 m x 0.2 m), spaced 8 cm apart to reduce inter-plant competition for nutrients. All plants were watered well daily until leachate application began on day 19 after seed germination. At this time, the plants had an average leaf length of 3.5 cm, 2.5 cm leaf width and 2.8 cm root length, as determined by destructive sampling. The experiment progressed for three consecutive growing seasons and the measured parameters were averaged over this period of analysis.

Nine treatments (Table 3) were imposed in a randomised complete block (RCB) design with stringent elimination of weeds. Each treatment contained two plants per bag with three replicates. Both 100% raw and treated leachate presented the same total N contents (0.090 N), whereas the 75–12.5% diluted treated leachate (DTL) contained 0.068–0.012% N. The N content of 50% DTL and 100% inorganic fertiliser (IF) was standardised at 0.045% N. For each treatment, 200 mL of leachate was applied by dripping twice a day (before 8:00 and after 17:00) until the plants expanded. At this time, higher evapotranspiration occurred and thus fertigation was increased to 250 mL (twice a day) to ensure 70% soil water holding capacity. The total leachate treatment period was 36 days. No pesticide was applied to the plants until the end of the experiment.

After day 56 (which is equivalent to 36 consecutive days of leachate application), the plants were harvested by uprooting. Fresh weights of leaves, stems and roots were determined, as well as total leaf number. Leaf length (LL), leaf width (LW) and stem height (SH) were measured prior to harvesting and root length was measured immediately after harvest.

The leaves, stems, and roots were dried at 70 °C in a forced draft oven (GO-251, Durham Geo-Enterprises Inc. Stone Mountain, GA, USA) for 3 days until constant weight to determine dry matter yield. All data generated were subjected to statistical analysis with one-way analysis of variance (ANOVA) using SPSS 17.0. Specific growth rate for leaf length, leaf width root length, and stem height were determined using the equation stated by Dimitriou et al.[37]:

$$\text{Specific growth rate} = \frac{\ln L_2 - \ln L_1}{t_2 - t_1} \text{(mm/day)} \qquad \text{Equation 1}$$

where L_1 and L_2 are the initial and final lengths at the exponential phases, respectively and t_1 and t_2 are the initial and final selected time intervals, respectively. Data were analysed using ANOVA and Microsoft Excel.

Heavy metal analysis

After drying, the weight of the plant and soil samples was recorded to the nearest gram. The samples were individually mashed to pass through the 2 mm screen in a laboratory mill (Serial no. 39017, Christy and Norris LTD, Chelmsford, England) and then 1 g of the sample was accurately weighed and placed into a 500 mL volumetric flask. For the plant samples, 1 mL of 65% HNO_3 and 10 mL of distilled water were added into the 500 mL volumetric flask and refluxed for 10 min by mounting the flask on a digestion heater (EAM9203 heating mantle, Camlab Ltd., Cambridge, UK) at 105 °C. Another 5 mL of 65% HNO_3 was added after 15 min and the mixture was allowed to digest until the solution became transparent. For the soil samples, 3 mL of (30%) H_2O_2 and 10 mL of HCl were added into the 50 mL volumetric flask while refluxing for 15 min. The resulting solutions were allowed to cool before being filtered and diluted with 50 mL of deionised water. The solutions were analysed for K, Ca, Mg, Na, Pb, Cd, Se, Al, Mn, Cu, Zn, Fe and As through inductively coupled plasma optical emission spectrometry.

Table 2: Physicochemical characterisation of the Mollisol soil prior to fertigation

Parameters	Soil	MAFF standard [(mg/kg)[44,45]	Methods
Clay (%)	25	–	Laboratory method (PSA)
Silt (%)	55	–	Laboratory method (PSA)
Sand (%)	20	–	Laboratory method (PSA)
Moisture (%)	40	–	Gravimetric analysis
pH	6.02	–	APHA 4500-H^+B
Total Nitrogen (%)	1.10	–	Kjeldahl method
Phosphorus (%)	21.8	–	USEPA 3050B
CEC (meq/100g)	36.90	–	Protocol HRN ISO 11260:2004
K (mg/kg)	9.11	–	USEPA 3050B
Ca (mg/kg)	44.73	–	USEPA 3050B
Mg (mg/kg)	5.83	–	USEPA 3050B
Na (mg/kg)	4.68	–	USEPA 3050B
Pb (mg/kg)	>0.06	0.19	USEPA 3050B
Cd (mg/kg)	>0.01	0.07	USEPA 3050B
Se (mg/kg)	>0.05	–	USEPA 3050B
Al (mg/kg)	1.85	–	USEPA 3050B
Mn (mg/kg)	8.85	–	USEPA 3050B
Cu (mg/kg)	2.71	2.9	USEPA 3050B
Zn (mg/kg)	0.65	0.16	USEPA 3050B
Fe (mg/kg)	30.47	–	USEPA 3050B
As (mg/kg)	>0.01	0.16	USEPA 3050B

PSA, particle size analysis

Table 3: Leachate treatment applications used the in experiment

Treatments	Details of Treatments
100%RL	100% Raw Leachate
100%TL	100% Treated Leachate
75% DTL	75% Treated Leachate +25% distilled water
50%DTL	50% Treated Leachate + 50% distilled water
50%DTL+50%IF	50%Treated Leachate + 50% Inorganic Fertilizer
25% DTL	25% Treated Leachate + 75% distilled water
12.5%DTL	12.5% Treated Leachate + 87.5% distilled water
100%IF	Inorganic Fertilizer N: P: K + 100 % distilled water
dH_2Ocontrol	100% Distilled Water (Control)

DTL, diluted treated leachate; TL, treated leachate; IF, inorganic fertiliser; RL, raw leachate

Results and discussion

Overview of leachate characteristics for fertigation

The total N content of the raw leachate sample was 900 mg/L (Table 1). The high N concentration can be attributed to the breakdown of nitrogenous substances during organic waste decomposition.[38] The Amper Tenang landfill leachate exhibited typical characteristics of an ageing methanogenic landfill with a BOD_5/COD ratio between 0.06 and 0.08 and pH within the range of 6.12–7.04. Christensen et al.[39] described this as a characteristic of an ageing landfill. The high suspended solid level in the leachate may be attributed to the presence of organic and inorganic compounds.[17]

After leachate treatment with 4 g/L $FeCl_3$ at pH 7, the contents of Cd, Al, Fe, Pb, Cu and Zn decreased by 100%, 64.4%, 51.9%, 82%, 56.8% and 96.6%, respectively. The optimal removal capacity for suspended solids (SS) was 80%. Hamidi et al.[40] reported the same dosages with similar reduction effects on colour, turbidity, and SS.

Plant physical growth evaluation

B. rapa L. survived until harvest, although common symptoms of soil salinisation such as chlorosis and leaf burn were noticed on 100%RL to 100%TL. This finding was not observed in plants irrigated with 75%DTL–12.5%DTL, probably because of the changes and/or decrease in concentration gradient. Plants receiving 25%DTL produced significantly longer leaves (23.17±0.577 cm) than those in the other treatments ($p<0.05$) (Table 4). Plants treated with 25%DTL presented wider leaves, with 1.36 and 3.23 times higher width than plants receiving inorganic fertiliser [100% IF (N_{15}:P_{15}:K_{15})] and the control. Less expanded leaf length and width were also observed in plants treated with the control compared with those in other treatments, which may be a clear indication of nutrient deficiency and differential N proportions.[41,37] Overall, 25%DTL could be the optimal nutrient requirement level for leaf expansion of *B. rapa* L. (Table 5), at a specific growth rate of 0.53 mm/day.

Table 4: Comparison of leaf length, leaf width, stem height and total leaf number for *Brassica rapa* L. after 56 days

Treatments	Leaf length (cm)	Leaf width (cm)	Stem height (cm)	Total leaf number*(not in cm)
100%RL	19.17±0.577^{zcfg}	9.00±0.866^{z}	2.44±0.323^{z}	16.17±1.893^{zdeh}
100%TL	19.50±1.323^{zcfgh}	9.83±1.155^{zh}	2.23±0.322^{x}	18.67±1.610^{zcdefh}
75%DTL	15.72±1.114^{z}	8.33±1.258^{z}	2.43±0.416^{z}	14.33±0.764^{z}
50%DTL	19.00±1.803^{zcf}	9.00±0.500^{z}	2.27±0.306^{x}	12.00±1.323^{z}
25%DTL	23.17±0.577^{zybcdfgh}	10.33±0.289^{zch}	2.30±0.361^{x}	12.33±1.041^{z}
12.5%DTL	15.67±1.041^{z}	9.33±1.443^{z}	2.17±0.379^{x}	15.17±2.363^{zde}
50%DTL+50%IF	16.83±2.021^{z}	9.17±1.155^{z}	2.50±0.866^{x}	16.33±1.607^{zdeh}
100%IF	17.00±1.500^{z}	7.67±0.764^{z}	2.50±0.866^{x}	13.23±1.662^{z}
dH_2O	7.17±0.577	4.50±1.803	1.50±0.500	8.33±0.764
TOTAL	**153.23±16.533**	**77.16±18.466**	**20.33±4.339**	**126.56±13.027**
Levels of significance:	**p<0.05 at F=35.256**	**p<0.05 at F=7.035**	**p<0.05 at F=1.03**	**p<0.05 at F=11.651**

DTL, diluted treated leachate; TL, treated leachate; IF, inorganic fertiliser; RL, raw leachate

Letters indicate statistical significance between different treatment levels using analysis of variance (ANOVA) version SPSS 17.0.

Table 5: Specific growth rate comparison for *Brassica rapa* L. at harvest after 56 days

Treatments	Leaf length (mm/day)	Leaf width (mm/day)	Stem height (mm/day)	Root length (mm/day^{-day})
100%RL	0.47	0.35	0.22	0.30
100%TL	0.48	0.38	0.22	0.42
75%DTL	0.42	0.33	0.22	0.40
50%DTL	0.46	0.35	0.20	0.38
25%DTL	0.53	0.39	0.21	0.30
12.5%DTL	0.42	0.37	0.19	0.40
50%DTL+50%IF	0.43	0.36	0.23	0.35
100%IF	0.44	0.31	0.23	0.42
dH_2O	0.20	0.16	0.09	0.21
TOTAL	**3.87**	**3.02**	**1.79**	**3.18**

DTL, diluted treated leachate; TL, treated leachate; IF, inorganic fertiliser; RL, raw leachate

Dry biomass weight evaluation

B. rapa L. that received 50%DTL+50%IF presented 2.25 and 1.60 times higher dry weight of leaf biomass than plants treated with the control and 100%IF (Table 6). This phenomenon may be attributed to the synergistic effect of N, which enhanced moisture retention in plants treated with 50%DTL+50%IF plants.[42] Statistical comparison showed that leaf dry weights were not significantly different for plants treated with 50%DTL+50%IF and those treated with 25%DTL $p<0.05$, which could be because of equal moisture contents and/or evapotranspiration rates during fertigation.[43] The dry root biomass of plants treated with 25%DTL was 3.16 and 1.70 times higher than that of plants treated with the control and 100%IF, respectively (Table 6). Initial low biomass yield, moisture content in the control and optimum growth nutrient requirements may be the implicating and/or limiting factors. Dry stem biomass was 5.48 times lower in control plants (0.25±0.050 g) than that in plants with 100%RL (1.37±0.176 g, $p<0.05$). The zero N supplementation during fertigation hindered the yield and subsequent biomass.

Heavy metal analysis in soil and plants, pH impact and N dynamics

The overall concentrations of heavy metals present in the soil prior to fertigation were lower than the detection and maximum permissible limits.[44,45] Application of leachate generally altered the physicochemical characteristics of the soil and the heavy metal uptake by vegetables.[46,6] A comparison between edible (plant) parts, namely, shoots collected 1 cm above the soil surface, showed that plants treated with 50%DTL and 100%IF and market samples of *B. rapa* L. (as control), contained zero Cd (Tables 7 and 8), while the permissible limit for Cd is 0.2 mg/kg.[47,48]

Table 6: Dry leaf, root and stem weights of *Brassica rapa* L. after 56 days

Treatments	Leaf (dry wt.) g	Root (dry wt.) g	Stem (dry wt.) g
100%RL	2.80±0.427^{zcd}	91±0.553^{zcdfgh}	1.37±0.176^{zcdefgh}
100%TL	2.55±0.247^{zd}	3.30±0.500^{zcdfgh}	1.21±0.110^{zdefh}
75%DTL	2.22±0.475^{x}	1.18±0.457^{x}	1.00±0.474^{zdefh}
50%DTL	1.95±0.377^{x}	0.95±0.182^{x}	0.33±0.11^{x}
25%DTL	3.60±0.304^{zybcdfh}	3.60±0.654^{zycdfgh}	021±0.015^{x}
12.5%DTL	2.67±0.301^{zd}	1.25±0.050^{x}	0.58±0.104^{e}
50%DTL+50%IF	3.95±0.050^{zybcdfh}	1.33±0.104^{x}	0.93±0.104^{zdefh}
100%IF	2.48±0.225^{z}	2.11±0.202^{zcdfg}	0.33±0.029^{x}
dH_2O	1.75±0.180	1.14±0.051	0.25±0.050
TOTAL	**23.97±2.586**	**17.77±2.753**	**6.21±1.178**
Level of significance:	$p>0.05$ at F=15.876	$p>0.05$ at F=24.543	$p>0.05$ at F=17.25

DTL, diluted treated leachate: TL, treated leachate: IF, inorganic fertiliser: RL, raw leachate

Letters indicate statistical significance between different treatment levels using analysis of variance (ANOVA) version SPSS 17.0.

Table 7: Metal content comparisons of 50% diluted treated leachate (DTL) with 100% IF treatment levels with both (water and market) controls for *Brassica. rapa* L. after 56 days

Metals	50%DTL (mg/kg)			100%IF (mg/kg)			dH_2O Control (mg/kg)			Market Vegetable Control (mg/kg)		
	Shoot	Root	Soil	Shoot	Root	Soil	Shoot	Roots	Soil	Shoot	Root	Soil
K	42.84±4.1	3.16±4.3	1.21±0.0	62.89±3.9	3.27±2.0	1.31±0.2	0.95±0.3	1.84±0.2	1.31±0.0	12.46±2.2	11.24±3.1	–
Ca	18.35±2.9	14.06±3.7	5.32±0.1	41.99±4.0	2.56±0.9	6.06±0.2	0.22±0.1	1.26±0.1	5.7±0.4	6.6±1.2	3.43±0.8	–
Mg	5.65±0.03	4.85±0.5	0.92±0.0	5.74±1.4	4.96±1.2	1.83±0.1	3.06±0.8	5.73±0.5	2.04±0.0	1.78±0.2	10.85±0.2	–
Na	41±02±3.2	30.40±4.4	5.29±0.1	15.53±1.4	2.93±1.9	6.03±0.1	2.16±1.0	2.93±0.2	2.97±0.0	5.26±2.1	3.65±0.1	–
Pb	0.09±0.12	0.41±0.3	0.00±0.0	0.07±0.10	0.19±0.1	0.00±0.0	0.00	0.00	0.00	0.00	0.00	–
Cd	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	–
Se	1.36±0.10	1.97±0.2	1.50±0.0	0.12±1.2	1.24±0.0	1.35±0.3	0.04±0.0	0.04±0.0	1.23±0.0	1.15±0.0	1.09±0.0	–
Al	20.05±3.5	16.11±2.7	3.12±0.0	0.83±2.3	3.49±1.6	2.62±0.0	4.55±0.7	4.80±0.8	5.69±0.2	4.55±0.0	4.80±0.8	–
Mn	0.43±0.22	0.49±0.1	0.07±0.0	0.18±0.4	0.43±0.0	0.13±0.0	0.17±0.1	0.08±0.0	0.15±0.0	1.22±0.2	0.20±0.0	–
Cu	0.16±0.02	0.25±0.0	0.07±0.01	0.01±0.36	0.20±0.2	0.02±0.0	0.04±0.0	0.03±0.0	0.09±0.0	0.10±0.0	0.07±0.0	–
Zn	0.82±0.47	1.16±0.1	0.32±0.00	0.23±1.56	1.29±0.4	0.40±0.0	0.40±0.1	0.33±0.1	1.00±0.0	0.90±0.1	0.34±0.1	–
Fe	8.82±4.15	9.98±0.4	0.37±0.00	0.91±2.55	2.51±2.8	0.28±0.0	4.34±0.4	4.00±0.4	0.93±0.0	4.11±1.1	6.50±1.2	–
As	0.08±0.03	0.08±0.0	0.12±`0.00	0.03±0.08	0.06±0.0	0.12±0.0	0.07±0.0	0.08±0.0	0.17±0.0	0.17±0.0	0.09±0.0	–

Table 8: Food and Agricultural Organization of the United Nations/World Health Organization (FAO /WHO) heavy metal permissible limits in vegetables

Metals	FAO/WHO (mg/kg)
K	N.A
Ca	75
Mg	N.A.
Na	N.A.
Pb	3
Cd	0.2
Se	N.A.
Al	N.A.
Mn	0.2
Cu	40
Zn	60
Fe	N.A.
As	1.0

N.A, not available

Source: FAO/WHO (Codex Alimentarius[47,48]).

Traces of Pb ranging from 0.07–0.09 mg/kg were detected in *B. rapa* L. edible parts treated with 50%DTL and 100%IF, whereas zero Pb content was observed in the market sample. Nevertheless, Pb concentration was still lower than the maximum permissible concentration of 3 mg/kg. Moreover, arsenic content was lower than the maximum permissible concentration limit of 1.0 mg/kg, as proposed by the Food Quality and Standard Control Division, Ministry of Health Malaysia[49] for heavy metals under the Malaysia Standard.

Large amounts of K, Ca, Mg, Na, Al, and Fe accumulated in the edible parts of *B.rapa* L. treated with 50%DTL were compared with those in the market sample, but the concentration of these elements was still within the acceptable range. This result is in agreement with a previous study by Liu et al.[50] that reported increased levels of metals in edible parts of food crops that were continuously irrigated with wastewater. Conventional wastewater treatment processes concentrate sludge fractions with accumulated heavy metals, thereby producing pure water with minimal metal concentrations.[51] The results of this study were obtained solely from a greenhouse experiment and should therefore be verified with field studies before being advocated to farmers. Long-term fertigation with landfill leachates may improve soil health conditions because of the fertiliser effect and provide economic benefits to farmers. However, this process may present certain limitations, particularly in terms of soil salinisation, N oversupply, heavy metal leaching to ground water and food chain contamination.[51]

The soil pH was 6.02±0.01 before fertigation began and then decreased to 6.0±0.001 and 6.01±0.002 for the undiluted (100%) and diluted (12.5–75%) treatments, respectively. Cation exchange capacity, clay fraction, and soil organic matter were implicated in influencing the buffering system in acid soils, as reported by Jiang et al.[52] N incorporation from leachate was assumed to be responsible for the decrease in pH. A field inquiry performed by Van Breemen et al.[53] and Huang et al.[54] indicated that 500 kg N/ha per year as urea is commonly applied the for production of *B. rapa* L. Based on this urea content, which is equivalent to 35.7 kmol N/ha per year, the estimated N input to 5 kg of Mollisol in a poly bag is 0.14 kmol N kg/ha per year. As the N concentration of undiluted leachate (100%) was 0.09% and that of diluted leachate (12.5–75%) ranged from 0.012–0.068%, the increase in H^+ protons$_{(pro)}$ caused by nitrification (Ni), amounted to 0.013 and 0.002 to 0.009 kmol kg/ha per year respectively, as calculated using Equation 2[55]:

$$H^+_{(pro)}\ (Ni) = 0.14 \times \%\ N_{(Leachate)}\ \text{kmol kg/ha per year} \qquad \text{Equation 2}$$

Thus, N use efficiency from harvested plants in the undiluted (100%) and diluted (12.5–75%) leachate treatments amounted to 40% and 60%, respectively. Hence, H^+ proton deposition$_{(dep)}$ in soil calculated using Equation 3,[56] was 0.06 and 0.08 kmol kg/ha per year for the undiluted diluted leachates, respectively.

$$H^+_{(dep)}\ (\text{uptake}) = 0.14 \times \%N_{(utilised)} \qquad \text{Equation 3}$$

The duration (1/T) for a unit decrease in soil pH on account of leachate treatments was estimated with Equation 4:

$$1/T = H^+_{(dep)} \times \text{Bulk Density [BD = 0.02: mass (5 kg)/vol. (250mL)]} \qquad \text{Equation 4}$$

For the undiluted (100%) and diluted (12.5–75%) leachates, the estimated duration was 625 years and 833 years, respectively. The soil acidification rate induced by N fertigation with leachate was approximately 0.01 unit pH/year.

Conditions such as extreme biogeochemical/anthropogenic disruptions in the soil ecosystem may alter and/or decrease soil pH and enhance cation bioavailability and desorption from soil matrices. In cases with rapid changes in soil pH because of fertigation with leachate, application should be terminated and waste composition from the leachate source, treatment facility, and concentrations applied to soil should be re-evaluated.

Generally, *B. rapa* L. exhibits higher mineral accumulation tendencies at their leaf region. This study revealed that K, Ca, Mg, Na, Al, and Fe were the most dominant minerals present in the plant. Pb and Cd concentrations were lower than the permissible levels.

Conclusion

This study attested that treated landfill leachate, similar to inorganic fertiliser, can be an effective source of nutrients for irrigated *B. rapa* L. Leachates can be recycled and utilised as bio-fertiliser for edible and non-edible plants, even for ornamental and timber species. The heavy metal levels in the treated leachate-grown *B. rapa* L. were lower than the stipulated permissible concentrations based on the FAO/WHO standards. Therefore, this treatment strategy will reduce the impact of chemical fertilisers on our ecosystem.

Acknowledgements

This research was supported by the IPPP-University of Malaya, Malaysia under grant No. IPPP/PV054/2011A. We thank the Solid Waste Laboratory, Institute of Graduate Studies, University of Malaya and the graduate students who supported this research work.

Authors' Contributions

F.O.A. was the principal researcher, while the P.A. was the project head and investigator. Both authors collaborated effectively to complete the study.

References

1. Dillion J, Edinger–Marshall S, Letey J. Farmers adopt new irrigation and fertilizer techniques. Calif Agr. 1999;53:24–28. http://dx.doi.org/10.3733/ca.v053n01p24
2. Keletso CM, Alphonsus KB, Felix DD. Symbiotic N nutrition, C assimilation, and plant water use efficiency in Bambara groundnut (*Vigna subterranea* L. Verdc) grown in farmers' fields in South Africa, measured using ^{15}N and ^{13}C natural abundance. Biol Fert Soils. 2014;50:307–319. http://dx.doi.org/10.1007/s00374-013-0841-3
3. Suja G, Sreekumar J. Implications of organic management on yield, tuber quality and soil health in yams in the humid tropics. Int J Plant Prod. 2014;8(3):291–310.
4. Swarnam TP, Vel Murugan A. Evaluation of coconut wastes as an amendment to acidic soils in low-input agricultural system. Commun Soil Sci Plan. 2014;45(8):1071–1082. http://dx.doi.org/10.1080/00103624.2013.867059

5. Cui Y, Li Y, Lu G, Sha Z. Nitrogen movement and transformation with different water supply for paddy rice. Adv Wat Sci. 2004;15:280–285.

6. Anita S, Rajesh KS, Madhoolika A, Fiona MM. Risk assessment of heavy metal toxicity through contaminated vegetables from waste water irrigated area of Varanasi, India. Int J Trop Ecol. 2010;51(2S):375–387.

7. Eriyamremu GE, Asagba SO, Akpoborie A, Ojeaburu SI. Evaluation of lead and cadmium levels in some commonly consumed vegetables in the Niger-Delta oil area of Nigeria. B Environ Contam Tox. 2005;75:278–283. http://dx.doi.org/10.1007/s00128-005-0749-1

8. Muchuweti M, Birkett JW, Chinyanga E, Zvauya R, Scrimshaw MD, Lester JN. Heavy metal content of vegetables irrigated with mixture of waste water and sewage sludge in Zimbabwe: Implications for human health. Agr Ecosyst Environ. 2006;112:41–48. http://dx.doi.org/10.1016/j.agee.2005.04.028

9. Satarug S, Haswell-Elkins MR, Moore MR. Safe levels of cadmium intake to prevent renal toxicity of human subjects. Brit J Nutr. 2000;84:791–802.

10. Marshall FM, Holden J, Ghose C, Chisala B, Kapungwe E, Volk J, Agrawal M, Sharma RK, Singh RP. Contaminated irrigation water and food safety for the urban and periurban poor: Appropriate measures for monitoring and control from field research in India and Zambia. Inception Report: DFID Enkar R8160. Brighton: SPRU, University of Sussex; 2007.

11. Sharma RK, Agrawal M, Marshall FM. Heavy metals contamination in vegetables grown in waste water irrigated areas of Varanasi, India. B Environ Contam Tox. 2006;77:311–318. http://dx.doi.org/10.1007/s00128-006-1065-0

12. Sharma RK, Agrawal M, Marshall FM. Heavy metals contamination of soil and vegetables in suburban areas of Varanasi, India. Ecotox Environ Safe. 2007;66:258–266. http://dx.doi.org/10.1016/j.ecoenv.2005.11.007

13. Vusumzi P, Ewa C, Luke C. Metal and flavonol contents of *Moringa oleifera* grown in South Africa. S Afr J Sci. 2013;109(3/4), Art. #835, 7 pages. http://dx.doi.org/10.1590/sajs.2013/835

14. Frank AS, Kees WV, Piet FO. A tiered approach for the human health risk assessment for consumption of vegetables from with cadmium-contaminated land in urban areas. Env Res. 2013;126:223–231. http://dx.doi.org/10.1016/j.envres.2013.08.010

15. Intawongse M, Dean JR. Uptake of heavy metals by vegetable plants grown on contaminated soil and their bioavailability in the human gastrointestinal tract. Food Addit Contam. 2006;23:36–48. http://dx.doi.org/10.1080/02652030500387554

16. Agamuthu P. Characterization of municipal solid waste and leachate from selected landfills in Malaysia. Malays J Sci. 1999;18:99–103.

17. Aziz HA, Daud Z, Adlan MN, Hung YT. The use of poly aluminum chloride for removing colour, COD and ammonia from semi-aerobic leachate. Int J Environ Eng. 2009;1:20–35. http://dx.doi.org/10.1504/IJEE.2009.026440

18. Öman CB, Junestedt C. Chemical characterization of landfill leachates – 400 parameters and compounds. Waste Manage. 2008;28:1876–1891. http://dx.doi.org/10.1016/j.wasman.2007.06.018

19. Santos DV, Sousa PL, Smith RE. Model simulation of water and nitrate movement in a level-basin under fertigation treatments. Agri Water Manage. 1997;32:293–306. http://dx.doi.org/10.1016/S0378-3774(96)01273-5

20. Guo D, Xiong Q, Xie C, Feng Y. Study on the relationship between the transforming of NO_3−N in soil and soil moisture under irrigation. Irrig Drain. 2001;20:66–70.

21. Ju XT, Liu XJ, Zhang FS. 2003. [Accumulation and movement of NO3-N in soil profile in winter wheat/summer maize rotation system.] Acta Pedo Sin. 40:538546. [Chinese].

22. Cheng CY, Chu LM. Phytotoxicity data safeguard the performance of the recipient plants in leachate irrigation. Environ Pollut. 2007;145:195–202.

23. Menser HA, Infant WM, Benneth OL. Spray irrigation with landfill leachate. Biocycle. 1983;24(3):22–25.

24. Liang J, Zhang J, Wong MH. Landfill leachate used as irrigation water on landfill sites during dry seasons. In: Wong MH, Wong JWC, Baker AJM, editors. Remediation and management of degraded lands. Boco Raton, FL: CRC Press LLC; 1999.

25. Cureton PM, Groenevelt PH, McBride RA. Landfill leachate recirculation: Effects on vegetation vigor and clay surface cover infiltration. J Environ Qual. 1991;20:17–24. http://dx.doi.org/10.2134/jeq1991.00472425002000010005x

26. Mapanda F, Managwayana EN, Nyamangara J, Giller KE. The effects of long-term irrigation using wastewater on heavy metal contents of soils under vegetable in Harare, Zimbabwe. Agr Ecosyst Envon. 2005;107:151–123. http://dx.doi.org/10.1016/j.agee.2004.11.005

27. Alados CL, Navarro T, Escòs J, Cabezudo B, Emlem JM. Translational and fluctuating asymmetry as tools to detect stress in stress-adapted and non-adapted plants. Int J Plant Sci. 2001;162(3):607–616. http://dx.doi.org/10.1086/320130

28. Zvera EL, Kozlov MV, Haukioja E. Stress responses of *Salix borealis* to pollution and defoliation. J Appl Ecol. 1997;34(6):1387–1396. http://dx.doi.org/10.2307/2405256

29. Mangwayana ES. Heavy metals pollution from sewage sludge and effluent of soil and grasses at Crowborough Farm [BSc thesis]. Harare: University of Zimbabwe; 1995. p. 63.

30. Nyamangara J, Mzezewa J. The effect of long-term sewage sludge application on Zn, Cu, Ni and Pb levels in a clay loam soil under pasture grass in Zimbabwe. Agr Ecosyst Envon. 1999;73:99–204.

31. Huxley A. The new RHS dictionary of gardening. New York: MacMillan Press; 1992.

32. Larkcom J. Oriental Vegetables. London: John Murray; 1991.

33. Grinsted MJ, Hedley MJ, White RE, Nye PH. Plant-induced changes in the rhizosphere of rape (*Brassica napus* var. Emerald) seedlings. I. pH change and the increase in P concentration in the soil solution. New Phytol. 1982;91:19–29. http://dx.doi.org/10.1111/j.1469-8137.1982.tb03289.x

34. Robinson HD. Development of methanogenic condition within landfill. Second International Waste Management and Landfill Symposium; 1989 Oct 9–13; Porto Coute, Sardinia, Italy.

35. Said NA. Physicochemical treatment of Bukit Tagar sanitary landfill leachate using Praestol 189K and aluminum chloride [Masters dissertation]. Kuala Lumpur: University of Malaya. 2008. p. 120.

36. United States Department of Agriculture. Soil Taxonomy. A basic system of soil classification for making and interpreting soil surveys. 2nd Ed. Washington, DC: Natural Resources Conservation Service; 1999.

37. Dimitriou I, Aronsson P, Weih M. Stress tolerance of five willow clones after irrigation with different amounts of landfill leachate. Bioresource Technol. 2006;97:150–157. http://dx.doi.org/10.1016/j.biortech.2005.02.004

38. Peter K, Morton A, Balaz A, Rooker P, Anders B, Anna L, et al. Present and long term composition of MSW landfill leachate: A review. Crit Rev Env Sci Technol. 2002;32(4):297–336. http://dx.doi.org/10.1080/10643380290813462

39. Christensen TH, Jensen DL, Filip Z, Gron C, Christen TH. Characterization of the dissolved organic carbon in landfill polluted ground water. Water Res. 1999;23:125.

40. Hamidi AA, Salina A, Mohd NA, Faridah AH, Asaari MSZ. Colour removal from landfill leachate by coagulation and flocculation processes. Bioresource Technol. 2007;98:218–220. http://dx.doi.org/10.1016/j.biortech.2005.11.013

41. Aronsson P. Nitrogen retention in vegetation filters of short rotation willows coppice [PhD thesis]. Uppsala: Swedish University of Agricultural Sciences. 2000. p.161.

42. Zalesny JA, Zalesny Jr RS, Coyle DR, Hall RB. Growth and biomass of *Populus* irrigated with landfill leachate. Forest Ecol Manag. 2007;248:143–152. http://dx.doi.org/10.1016/j.foreco.2007.04.045

43. Montree I, Qifu M, David WT. Photosynthetic and growth responses of juvenile Chinese kale (*Brassica oleracea* var. *alboglabra*) and caisin (*Brassica rapa* sub sp. *parachinensis*) to water logging and water deficit. Scientia Hortic-Amsterdam. 2007;111:107–113. http://dx.doi.org/10.1016/j.scienta.2006.10.017

44. Ministry of Agriculture, Fisheries and Forests and Department of the Environment (MAFF). Review of the rules for sewage sludge application to agricultural land. Report of the Independence Scientific Committee: PB 1561. London: MAFF; 1993.

45. Ministry of Agriculture, Fisheries and Forests and Department of the Environment (MAFF). Survey of the cadmium contained in agricultural products. Japan: MAFF; 2002. p.1–45.

46. Monu A, Bala K, Shweta R, Anchal R, Barinder K, Neeraj M. Heavy metal accumulation in vegetables irrigated with water from different sources. Food Chem. 2008;111:811–815. http://dx.doi.org/10.1016/j.foodchem.2008.04.049

47. Codex Alimentarius. Maximum levels for cadmium in cereals, pulses and legumes. CAC/GL 39-2001. Rome: Joint Food and Agricultural Organization of the United Nations/World Health Organization; 2001.

48. Codex Alimentarius. Maximum levels for lead. Codex STAN 230-2001. Rome: Joint Food and Agricultural Organization of the United Nations/World Health Organization; 2001.

49. Ministry of Health Malaysia. Food safety and quality division [document on the Internet]. c2013 [cited 2014 Nov 18]. Available from http://fsq.moh.gov.my.

50. Liu WH, Zhoa JZ, Ouyang ZY, Soderlund L, Liu GH. Impacts of sewage irrigation on heavy metals distribution and contamination. Environ Int. 2005;31:805–812. http://dx.doi.org/10.1016/j.envint.2005.05.042

51. Weiping C, Sidan L, Wentao J, Meie W. Reclaimed water: A safe irrigation water source? Environ Dev. 2013;8:74–83. http://dx.doi.org/10.1016/j.envdev.2013.04.003

52. Jiang J, Xu RK, Zhao AZ. [Determination of pH buffer capacity of acid red soils by acid-base titration.] Chinese J Soil Sci. 2006;37:1247–1248. Chinese.

53. VanBreemen N, Mulder J, Driscoll C. Acidification and alkalinisation of soils. Plant Soil. 1983;75:283–308. http://dx.doi.org/10.1007/BF02369968

54. Huang P, Zhang J-B, Xin X-L, Zhu A-N, Zhang C-Z, Ma D-H, et al . Proton accumulation accelerated by heavy metal chemical nitrogen fertilization and its long-term impact on acidifying rate in a typical arable soil in the Huang–Huai Plain. J Integr Agr. 2015;14(1):148–157. http://dx.doi.org/10.1016/S2095-3119(14)60750-4

55. Xing GX, Zhu ZL. An assessment of N loss from agricultural fields to the environment in China. Nutrient Cycl Agroecosys. 2000;57:67–73. http://dx.doi.org/10.1023/A:1009717603427

56. Zhang YM, Chen DL, Zhang JB, Edis R, Hu CS, Zhu AN. Ammonia volatilization and denitrification losses from an irrigated maize-wheat rotation field in the North China plain. Commun Soil Sci Plan. 2004;35(19/20). http://dx.doi.org/10.1081/css-200036499

The evolving landscape of plant breeders' rights regarding wheat varieties in South Africa

AUTHORS:
Charity R. Nhemachena[1]
Frikkie G. Liebenberg[1]
Johann Kirsten[1]

AFFILIATION:
[1]Department of Agricultural Economics, Extension and Rural Development, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Charity Nhemachena

EMAIL:
cmudombi@yahoo.com

POSTAL ADDRESS:
Department of Agricultural Economics, Extension and Rural Development University of Pretoria, Private bag X20, Hatfield, Pretoria, 0028, South Africa

DATES:
Received: 04 May 2015
Revised: 13 Aug. 2015
Accepted: 05 Sep. 2015

KEYWORDS:
Plant variety; protection; breeding; varietal improvements; trends

HOW TO CITE:
Nhemachena CR, Liebenberg FG, Kirsten J. The evolving landscape of plant breeders' rights regarding wheat varieties in South Africa. S Afr J Sci. 2016;112(3/4), Art. #2015-0164, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20150164

Addressing the multiple challenges facing global agriculture requires integrated innovation in areas such as seeds, biotechnology, crop protection, grain storage and transport. Innovations related to plant improvement and the development of new or improved plant varieties will only happen at an optimal level if plant breeders' rights (PBR) are properly protected. The objective was to analyse the evolving landscape of wheat plant breeders' rights to address the dearth of empirical evidence of the patterns and trends of wheat varietal improvements in South Africa. We compiled a detailed and novel count and attribute database of wheat varietal innovations in South Africa from 1979 to 2013 using various sources. This data set was then analysed to ascertain the main trends in, and ownership of PBRs for wheat varietal improvements in South Africa over this period. A total of 134 PBR wheat varietal innovations were lodged from 1979 to 2013, an average of 6 applications per year. The administrative delays in granting PBR applications were substantially reduced by 77 days during the post-deregulation period (after 1996), indicating increased efficiency. The main PBR applicants were Sensako (39%), the Agricultural Research Council Small Grains Institute (ARC-SGI) (25%) and Pannar (15%). The ARC-SGI contributed to some of the PBRs owned by private companies through shared genetic resources before Plant Variety Protection (PVP) was implemented. Future innovations and dissemination of wheat innovations can be stimulated by plant variety protection, together with broader variety sector legislation that encourages both public and private sector investment.

Introduction and background

Integrated agricultural innovations in areas such as seeds, biotechnology, crop protection, agronomic management practices, harvesting and processing are critical to help address the global challenges facing agriculture, such as climate change and decreasing availability of water and farmland, increasing food demand resulting from increasing population size, and demand for food crops for other uses such as biofuels.[1,2] For example, varietal innovations that are high-yielding and drought-resistant are required if farmers are aiming to produce more with less inputs. However, investments in varietal improvements are expensive (in terms of skills, labour, material, resources etc.), take time to develop (10–15 years in the case of many plant species), and the resulting seed products face the risk of being reproduced and 'copied' by competitors, necessitating the need for some form of enforceable commercial protection for plant breeders.[1-3]

For a long time, developing countries have relied on public investments (national and international) for plant varietal improvements.[3,4] Most of the public research investments were implemented through collaborations between national agricultural research institutes and the International Agricultural Research Centres of the Consultative Group on International Agricultural Research such as The International Maize and Wheat Improvement Center (CIMMYT) in Mexico, and the International Rice Research Institute (IRRI) in Manila.[4] Despite the long history of public sector dominance in developing country research, there are growing calls for the private sector to play a major role in agricultural investments, including plant varietal improvements. This would help address the increasing funding challenges being faced by public research institutions. The private sector, on the other hand, requires economic incentives provided by Intellectual Property Rights (IPRs) to invest in plant variety innovations.[5] At the same time, the publicly funded research outputs should be protected, utilised and commercialised for the benefit of the funding country. All of this points to the importance of plant variety protection as being critical to stimulate plant breeding innovation and the dissemination of innovations.[6]

Pardey et al.[7,8] argue that plant varietal rights are subject to controversy and ongoing public policy scrutiny and debate. In order to inform these policy debates, there is a need to understand the evolution of varietal rights and the extent of the varietal rights which are granted. In addition, it is also important to understand the changes which are experienced over time regarding the rights on offer; changing ownership of the rights (including comparison between public and private, as well as domestic and foreign breeders); impacts of plant variety protection on varietal development etc.

In this research, the evolving landscape of plant breeders' rights was analysed regarding wheat varietal improvement in South Africa. The earlier efforts by Stander[9] to undertake a similar exercise were limited by the shorter period of relevant observations, compared to the current study. In addition, we do not know the implications of plant variety protection on public and private investments for wheat varietal improvements in South Africa, or whether the granting of plant breeders rights has enhanced, slowed or changed the nature of genetic improvement. The study therefore focused specifically on the following research questions:

- Which trends can be observed in the wheat varietal improvements subsector in South Africa?
- How have plant breeders' rights for wheat varietal improvements evolved?
- Who are the main plant breeders' rights holders in the wheat sector?

- What are the effects of these developments for diversity and competition in the South African wheat breeding industry?
- Which measures can be taken by the South African Government to restrict or reverse possible negative impacts of these trends in light of relevant policy objectives?

International experiences in plant variety protection

The protection of intellectual property rights of plant breeders was recognised from the 19th century onwards. The International Union for the Protection of New Varieties of Plants (UPOV), established in 1961, seeks to harmonise plant variety protection (PVP) laws and standards of protection across member countries.[10] Plant variety protection was almost exclusively the protection granted for plant-related innovations, until 25 years ago.[1] Plant variety protection, also called 'plant breeders' rights', is defined as an independent *sui generis* form of protection of new plant varieties for essential features (usually phenotypical).[2] Plant breeders' rights (PBRs) constitute a form of intellectual property right that grants the breeder of a new plant variety the exclusive rights to benefit from his or her variety. In addition, innovations in plant breeding are driven by the need to acquire and/or increase market share.[3] Plant breeders' rights provide legal protection for the plant breeder to exclude others from commercialising the protected variety for a specified number of years, usually five years. This gives protection against unauthorised copying (propagation) of the protected variety for commercial uses by farmers and competitors. At the end of the protection period, the breeder can issue licences to other breeders who might be interested in using any material of the variety in their breeding activities.

Countries are required by the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) of the World Trade Organisation to provide plant varietal protection 'either by patents or by an effective *sui generis* system or a combination thereof (TRIPS Article 27(3)(b))'[11]. In addition to the obligation under the TRIPS Agreement for countries to introduce plant variety protection systems, other benefits include providing a system of incentives for individuals and entities (state, private and foreign) engaged in plant breeding to increase the quantity or effectiveness of plant breeding. Plant variety protection contributes to: increasing genetic diversity; improving food security through production of high quality and higher yielding varieties; sustaining agriculture by promoting use of national resources and inputs; and protection of the environment through pest and disease resistant varieties that need fewer chemicals (pesticides).[2]

The standard argument for implementing PVP is that they would stimulate and drive investments in plant breeding research and development of the domestic seed sector.[11] Although PVP provide some incentives for increased plant breeding, evidence from some empirical studies show mixed results with some authors finding that it may not be strong enough to encourage plant breeding investments.[11-14] For example, Tripp et al.[11] studied the potential of PVP to provide incentives for plant breeding in developing countries (China, India, Colombia, Kenya and Uganda). Their findings showed that development of PVP systems in developing countries should be framed as part of a broader strategy for seed system development, as PVP may not be adequate to initiate commercial seed development. In a separate study, Alston and Venner[13] found that the introduction of the PVP Act of 1970 in the USA increased public investments only (with no change in private investments) and did not affect experimental and commercial wheat yields.

Srinivasan[5] explored the levels in the ownership of IPRs over plant varieties, based on data from 30 UPOV member-countries. He argues that the combination of ownership of plant variety rights and growing plant variety protection systems in the developing world would have significant influence on future plant innovations and distribution of market power between companies. This has implications for the structure of the domestic seed industries and gaining access to protected varieties and associated plant breeding technologies. Diez[15] analysed the impact of plant variety rights on public and private research sectors in Spain, focusing on the role of PVP in altering research incentives in the understanding of intersectoral differences. The study showed that there were positive incentives, particularly for the private sector, for increased investments in plant breeding driven by higher appropriability conditions.

Wright and Pardey[16] argue that with the exception of a few, many public agricultural research institutions in developing countries have made little progress in developing and commercialising their innovations. Related to this, Alston and Venner[13] state that despite plant variety protection providing some exclusive rights to the plant breeder, sometimes the high enforcement costs and the difficulties in monitoring areas planted to a crop, such as wheat, and in determining the source of seed, make it difficult to exclude farmers and seed companies from freely using protected varieties.

Plant variety protection in South Africa

South Africa became a member of UPOV in 1978. According to Moephuli et al.[17], the country did not accede to UPOV 1991. In 1976 South Africa enacted the *Plant Breeders' Rights Act (Act 15 of 1976)*. This was later amended in 1996 to conform to the constitution, as well as to align and comply with the UPOV 1991. Plant variety protection in South Africa is also guided by the *Plant Improvement Act 1976 (Act 53 of 1976)*. Granted PBRs are listed in a variety listing and are usually granted for 20 years for all crops and 25 years for trees and vines. However, the national authority may expropriate rights for national interest, such as food security needs. The granted PBRs automatically expire at the end of the prescribed period.[17]

The PBR provides the owner of the variety the opportunity to obtain financial reward from the investments put into breeding and development of the new variety in order to recover the costs incurred. Plant breeders' rights have evolved since the nineteen thirties, when crop improvements became an applied form of genetics practised by specialised institutions and seed companies.[18] The owner of a PBR has the privilege of a sole right period of 5 years and during this period the owner does not award licences, thus securing a chance to obtain return on investment for plant varietal improvements. This means that the breeder is the only one who can use the materials of the variety while it is protected. After the 5-year period, the owner of a PBR can allow the granting of licences through an attorney and the relevant contract should show how royalties are to be paid. These licences allow the owner to receive payment on any breeding material that is used by other breeders from the variety. In the case where the holder of the right refuses to issue licences to other persons who wish to use and market the material from the protected variety in the breeding programmes, the Registrar from the Department of Agriculture, Forestry and Fisheries (DAFF) may issue a compulsory licence (De Bruyn E 2013, oral communication, Jan 22.). After the expiry of the full period of the plant breeders' rights, the variety becomes public property and anyone may propagate and sell it.

The holder of a set of plant breeder's rights is required to maintain the reproductive material during the period of the rights. It is, however, not clear what happens to the said material after the period of the rights has expired. There is a test for the distinctness of the candidate variety: it should be compared with other varieties of the same kind of plant, of which their existence on that date is a matter of common knowledge. The holder is not entitled to exclusively make available plant material after the holder's rights have expired. This is a problem for other interested parties, as such material will then not be available for comparison purposes with other material that enters the market. DAFF does not have facilities to maintain all the plant material of varieties which have been granted plant breeders' rights, so they are sent to the gene bank at the Agricultural Research Council Small Grains Institute (ARC-SGI).

In South Africa, 60% of holders of PBRs on most crops are foreigners that are largely based in Europe and North America, and this asymmetry is not unique to South Africa as a developing country.[18] The plant breeders' rights in South Africa give the holder a limited exclusive right to the variety. However, traditional farming practices, such as re-sowing and saving protected varieties, may constitute infringement of that right. The *Plant Breeder's Right Act (Act No 15 of 1976)* allows farmers to use

(re-sow) protected material on his or her own holding. This expression of farmer rights is known as farmer privileges.[18]

Research methods and data

Plant breeders' rights (PBR) in South Africa were analysed to assess the sources of intellectual property rights in wheat varietal improvements in South Africa. The paper used secondary information collected from various sources on wheat varietal improvement policies, and changes in wheat plant breeders' rights. Other studies that have analysed changes in plant variety protection, focusing on trends and changes in plant variety protection policies, include Diez[15], Srinivasan[5], Louwaars et al[3] and Pardey et al.[7,8] In these studies, trends were analysed to understand the evolution of plant varietal protection. This paper applies the same approach in order to understand the changes that have shaped the South African wheat varietal improvement landscape to date, and to allow a comparison with other countries at a later stage.

A detailed and novel count, and attribute database of wheat varietal innovations in South Africa from 1979 to 2013 was compiled. Consultations were made with key informants from the Department of Agriculture Forestry and Fisheries, ARC and the South African National Library during development of the database. The database provides information to assess changing amounts and forms of wheat plant breeders' rights, and changes in the types and pool of the applicants receiving the rights. The database gathered data on applications, granting, and rejections and surrendering of the PBRs. Additional information captured in the database includes plant variety name, alias name, applicant name, applicant type, application date of PBR, grant date of PBR, withdrawal date of PBR, and date end of sole right of PBR, showing end of sole right month and date of termination of PBR.

Data were manually compiled on wheat variety rights from information on wheat plant breeders' rights applications and granting obtained from the *Plant Variety Journal* of the South African Department of Agriculture Forestry and Fisheries. Additional data were gathered from the South African National Library and the ARC. These different sources of data were used to backfill the data series to come up with a complete database. The data series included 134 applicants for wheat variety rights lodged between January 1979 and December 2013. The empirical analyses were based on descriptive statistics, trend analysis and graphical representation of trends and ownership of wheat varietal improvement PBRs.

Results and discussion

Trends in plant breeders' rights for wheat varietal improvements

Figure 1 shows the trends in annual applications for PBRs for wheat varietal improvements lodged in South Africa. A total of 134 PBRs were lodged for wheat varietal innovations from 1979 to 2013, an average of 6 applications per year. Data on the number of PBRs lodged for wheat varietal improvements since 1979 show a fluctuation trend with some years having high numbers of applications, and others recording very low applications. For example, 2012 received the highest number of PBRs for wheat varietal improvements, with 14 applications, while 2004 recorded only 1 application. A possible explanation for the uneven trend in applications could be that some breeders chose to apply for their new varieties to be included on the national variety list but did not apply for varietal protection. The variety would then be included on the national variety list so it could be known by millers and bakers, but unless the owner applied for PBRs, anyone could use any material on the variety without paying the owner for IPRs.

Outbreaks of pests or diseases can trigger more investments in wheat varietal improvements in some years, compared with others. For example, the Ug99 wheat stem rust race group from East Africa was detected in 1999 and since then, 11 pathotypes within the race group have been identified across Africa and a few neighbouring countries. In South Africa, 'four races in the Ug99 lineage have been detected – TTKSF (in 2000), TTKSP (in 2007), PTKST (in 2010) and TTKSF+ (in 2010)'[19]. In Addition, a new Ug99 variant (race TTKSF+) was identified in South Africa and Zimbabwe in 2012[19]. The identification of new variants of Ug99 in recent years, particularly from 2010, could have triggered increased investments in wheat varietal improvements, leading to high numbers of PBR applications in 2012. Generally, the main triggers for wheat varietal improvements in South Africa include the outbreak of pests and diseases, such as wheat rust, septoria and Russian wheat aphid, and different environmental and climatic conditions.

It is also important to understand the effects of the abolishment of the South African Wheat Board in 1996 and the establishment of the ARC-SGI in 1991 on the number of applications for PBRs for wheat varietal improvements. Analysis of the time periods before and after the abolishment of the Wheat Board shows that 77% of the applications were lodged between 1998 and 2013 (after the abolishment of the Wheat Board). The number of PBR applications increased faster after the abolishment of the Wheat Board, at an average of 6 applications per year, compared with 4 applications per year for the time period 1979–1997.

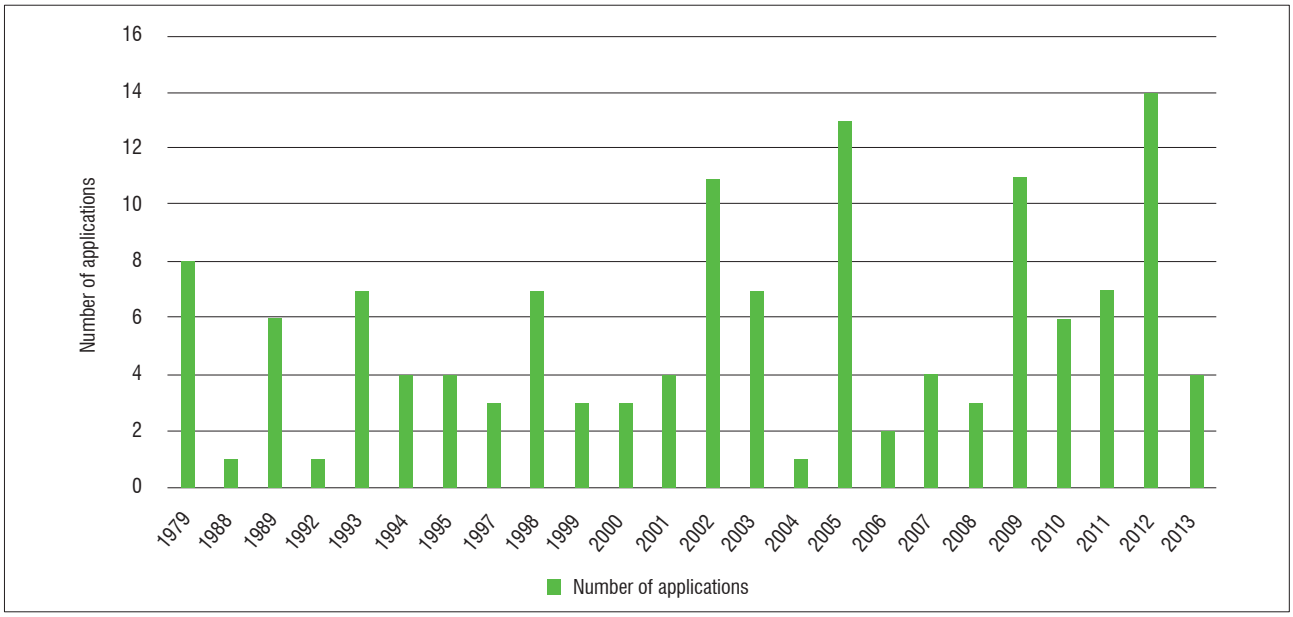


Figure 1: Annual applications for PBRs for wheat varietal improvements from 1979 to 2013, compiled from *Plant Variety Journals*.

In recent years, the numbers of wheat variety PBRs have substantially increased, compared with the past. From the results, the liberalisation of the wheat sector can be argued to have opened competition in wheat varietal improvements and stimulated increases in applications for PBRs on wheat varietal innovations. Further analysis by time period before and after the establishment of the ARC-SGI shows that only 15 applications for PBRs for wheat varietal innovations were lodged before 1991. The rest of the applications for wheat varietal innovation PBRs (89%) were lodged after the ARC-SGI was established, from 1992 to 2013. The rate of applications was six per year after the establishment of the ARC-SGI, compared with five applications per year before the ARC-SGI was established. The decreasing funding for agricultural research might be contributing to the ARC-SGI seeking revenue from their research outputs.

Figure 2 presents an analysis of the wheat varietal improvement PBRs applications, grants and lags. Panel (a) plots PBRs applications and grants in each year, while panel (b) tracks the lag in days between the application and grant dates of each PBR that was eventually granted a certificate. The average time of waiting between application and granting of protection was around 336 days for the overall period of study. Analysis of the grant lags by pre- and post-deregulation time periods indicates that before 1997 (pre-deregulation period) the grant lag was 451 days and this reduced to 374 days in the post-deregulation period. When the application process starts, the sample seeds and application papers are taken and inspected before being sent for trials at ARC. If the variety is found to be distinct, uniform and stable for two seasons, the variety is then granted the PBRs. This means that the processes would take at least 300 days for PBRs to be granted. If the trials are unstable, more trials may be required and the process would take longer until all the irregularities are checked and rectified (De Bruyn E 2015, Oral communication, Aug 24).

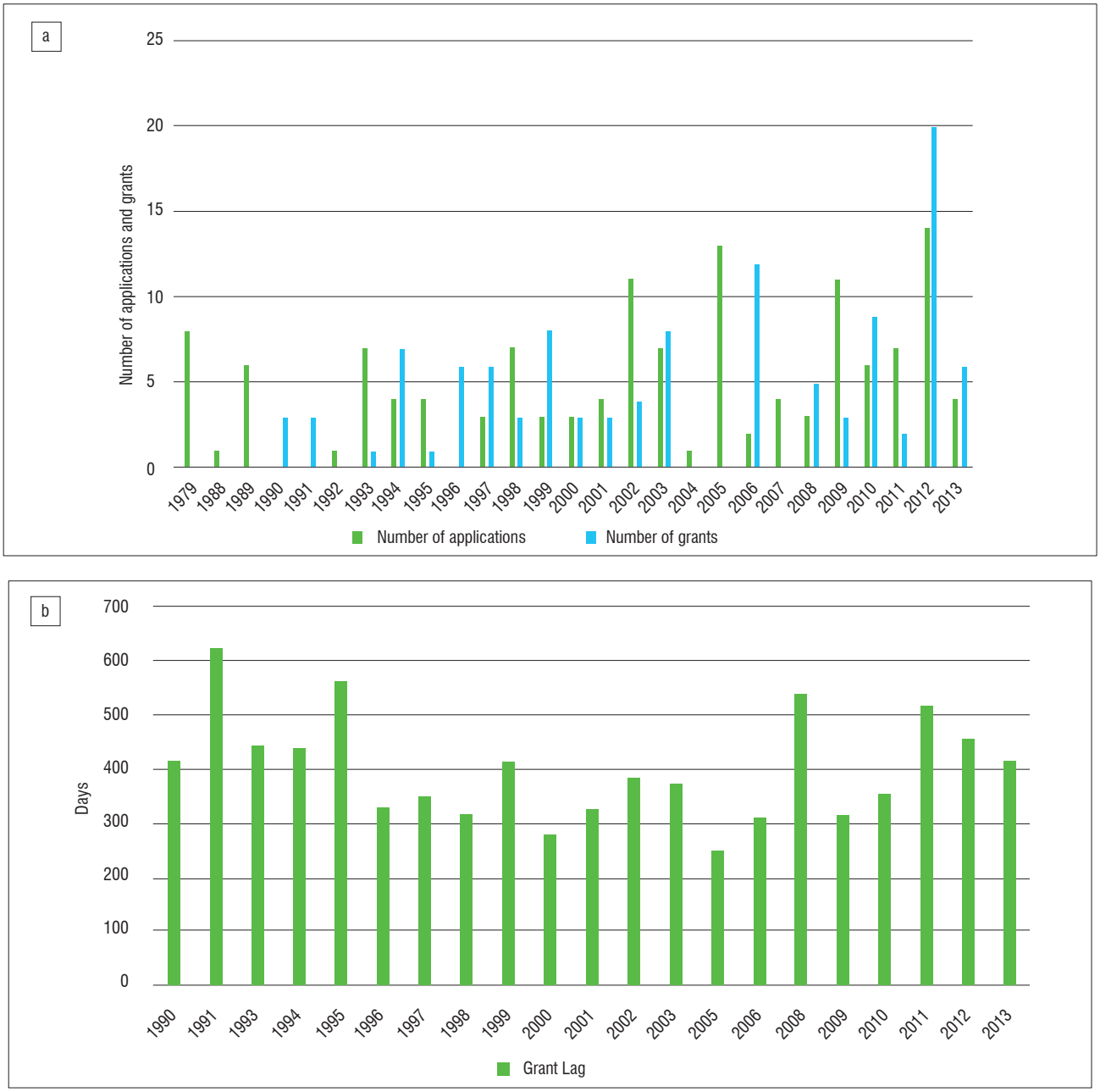


Figure 2: (a) Applications, grants and (b) grant lags in obtaining plant breeders' rights for wheat varietal improvements from 1979 to 2013 compiled from data in *South African Plant Variety Journals*.

Pardey et al[7] argued that administrative delays substantially lengthened the lags in the processing of applications. The grant lags (of about 300 days) were shorter after the abolishment of the Wheat Board in 1997, as well as after the establishment of the ARC-SGI. This can be considered to be efficient in comparison with the USA, where the average grant lag ranged from 500 days to 1449 days.

The plant variety rights applications can at any time be: granted; abandoned or withdrawn by the applicant; deemed ineligible – such as in cases of incomplete applications; denied by the office; or still pending examination.[7] In the South African context, applications for PBRs may be (a) granted if they meet the required standard by DAFF, (b) surrendered if there is need to do so if the applicant no longer needs protection and also after the completion of the right to be protected – 20 years in the case of wheat, (c) rejected if the application is incomplete and does not meet the required standards, and (d) terminated if the Registrar deems it necessary as in the event of certain irregularities (De Bruyn E 2013, oral communication, Jan 22).

Table 1: Changing disposition of plant breeders' rights applications *compiled from data in South African Plant Variety Journals*

Year	Number of wheat varietal PBR applications			
	Surrendered	Deleted	Rejected	Granted
1990	1	–	2	3
1991	2	2	–	3
1993	–	1	–	1
1994	4	3	–	7
1995	–	–	–	1
1996	3	2	–	6
1997	3	2	–	6
1998	1	1	–	3
1999	1	2	–	8
2000	1	–	–	3
2001	1	–	–	3
2002	1	1	–	4
2003	2	1	–	8
2006	3	–	–	12
2008	2	–	1	5
2009	1	–	–	3
2010	1	–	–	9
2011	–	–	–	2
2012	5	–	–	20
2013	1	–	–	6
Total	**33**	**15**	**3**	**113**

Table 1 shows the changing status of wheat varietal improvement PBRs, over time. For the period under review, 113 wheat variety rights applications were lodged between 1990 and 2013 and were granted, while 33 of the applications were surrendered, and 15 were deleted. The results show that only 3 cases were rejected from 1990 to 2013. Analysis of the granted PBRs for wheat varietal innovations by decade shows an average grant rate of 4 PBRs/year; 5 PBRs/year and 9 PBRs/year for the periods 1990–1999, 2000–2009 and 2010–2013, respectively. Further analysis by pre- and post-deregulation of the wheat sector shows that the grant rate increased from 4 grants/year before abolishment of the Wheat Board to an average of 7 grants/year from 1998 to 2013. In addition, from 2003 to 2013, the number of granted PBRs for wheat varietal improvements was relatively high, compared with other years, and the highest number was recorded in 2012 (20 PBRs were granted). The results indicate that plant breeders are increasingly seeking protection for their innovations.

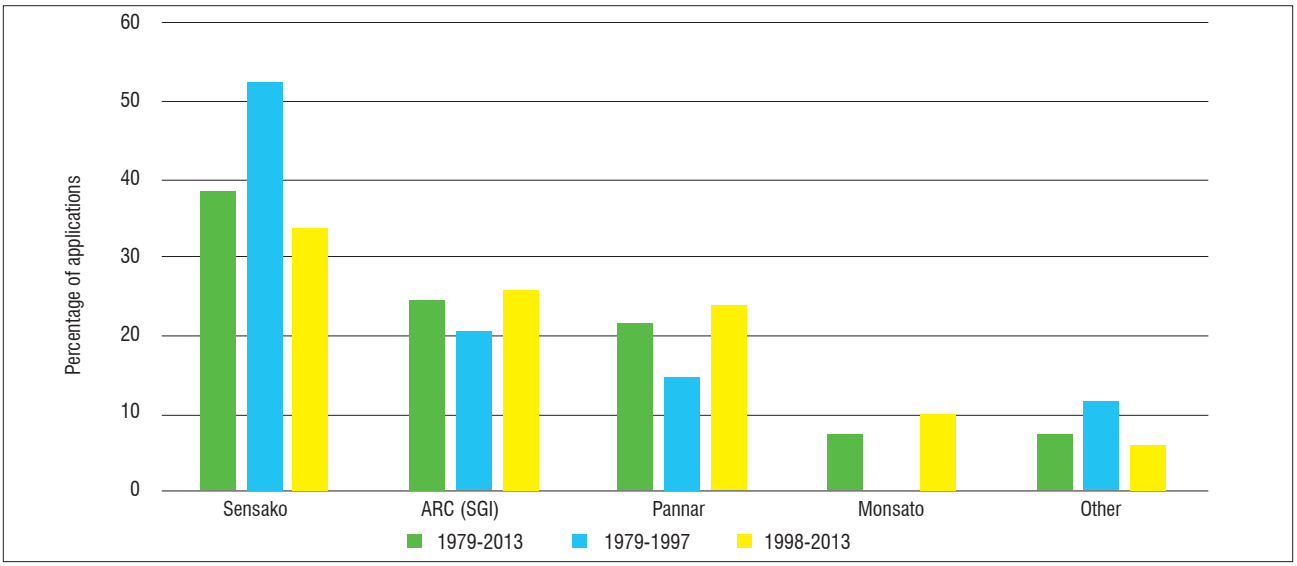
Applicants for wheat varietal plant breeders' rights

This section analyses the composition of applicants seeking PBRs for wheat varietal improvements from 1979 to 2013 with the focus on identifying the changing public and private roles among the applicants. Figure 3 presents the composition of applicants for PBRs for wheat varietal improvements from 1979 to 2013. Based on analysis of share of PBRs for wheat varietal improvements since the publication of the *South African Plant Variety Journal* in 1979, the main applicants were Sensako (39%), ARC-SGI (25%) and Pannar (15%). Before 1997, the share of applications from Sensako constituted 53% of all the applications, while that of the ARC-SGI was 21%, and Pannar had 15%. Other applicants, such as CIMMYT, University of Free State, Carnia Seeds, Cargill USA and Gwk Beperk, constituted 12% of the applications. Analysis of the period after the abolishment of the Wheat Board shows that Sensako's share decreased to 34%, while those of the ARC-SGI (26%) and Pannar (24%) increased by 5% and 9%, respectively. The share of applications and grants held by the ARC-SGI is comparable to other countries. In the Netherlands, Louwaars et al.[3] found that the public sector (universities, government bodies and private non-profit organisations) submitted 23.8% of plant-based patents, while in the USA, public bodies were granted 21.9% of the patents. Furthermore, private companies dominated the number of plant-based patents granted in the country.[3]

In many developing countries, national agricultural research systems, such as the ARC-SGI and Consultative Group on International Agricultural Research Institutions (CGIARs), dominate agricultural research, including varietal improvements. The development of new varieties is driven by the exchange of plant varieties and genetic resources between the national agricultural research systems and CGIAR institutions, unencumbered by IPRs.[5] In this case, ownership of IPRs was irrelevant to plant breeding and for accessing plant genetic resources from other countries.[16,5] Similarly, the ARC-SGI did not previously apply for PBRs, despite being actively involved with wheat varietal improvements. However, since the abolishment of the Wheat Board in 1997, the ARC-SGI increased their share of PBRs lodged by 5% to 26%, with most of the applications being made in recent years. It might be that the reduced funding for the ARC-SGI contributed to the institution seeking protection for its varieties as a way to generate additional revenue.

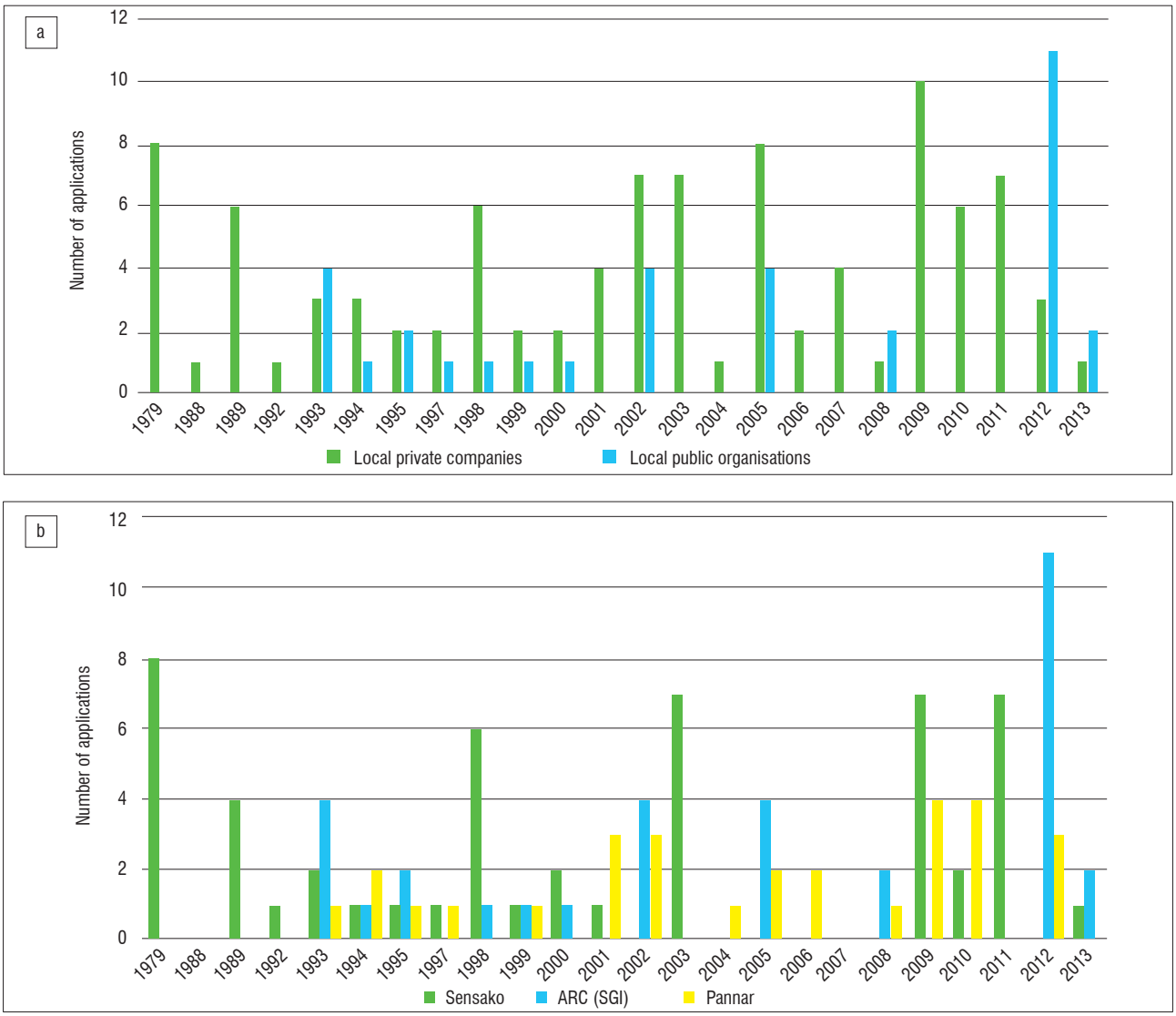
Figure 4 presents the changing composition of applicants for wheat PBRs. The majority of the applications for wheat varietal improvements were filed by private companies (Figure 4a), the two main actors being Sensako and Pannar (Figure 4b). Pardey et al.[8] found that the private sector constituted the largest share in the USA, accounting for 82% of total plant varietal rights, while the share of varietal rights owned by public entities (e.g. US and foreign universities, research foundations and government agencies) was very small, considering the high volume of research they perform. In the South African wheat sector, the public sector, especially the ARC SGI, continues to play a major role in the development of wheat varieties and has PBRs from wheat varietal improvements. Before the establishment of the ARC in 1991, the public sector made no applications for PBRs for wheat variety improvements, but since its establishment, the public sector applications for wheat variety PBRs drastically increased, with 4 applications in 1993, 4 in 2002, 2 in 2005 and 11 in 2012. The public sector made the largest number of wheat variety PBR applications (11) in 2012, compared with 3 from private companies in the same year.

It is important to note that public and private sector institutions freely exchanged plant genetic resources and breeding lines before the advent of IPRs.[5] Public sector institutions used to develop basic breeding lines and made them available for the development of 'finished products' by the private sector. The advent of PVP/IPRs considerably restricted the exchange of plant genetic resources and



ARC-SGI, Agricultural Research Council Small Grains Institute

Figure 3: The structure of applicants for plant breeders' rights for wheat from 1979 to 2013, compiled from data in *South African Plant Variety Journals*.



ARC-SGI, Agricultural Research Council Small Grains Institute

Figure 4: Applications for plant breeders' rights for (a) wheat by private vs public organisations and for (b) wheat made by the main role players from 1979 to 2013, compiled from data in *South African Plant Variety Journals*.

breeding lines between the public and private research institutions.[20] With this background, it can be argued that the ARC-SGI has largely contributed to some of the PBRs owned by private companies through shared genetic resources made available before the PVP/IPRs were implemented.

Implications of wheat plant variety protection trends for the Agricultural Research Council

PBRs are the predominant intellectual asset of the ARC.[17] Other forms of ARC intellectual assets include patents, copyrights (publications), trademarks, trade secrets and expertise. *The South African Intellectual Property Rights from Publicly Financed Research and Development Act No. 51 of 2008* prescribes that publicly funded research and development must be protected, utilised and commercialised. This is to ensure that: it benefits the people of South Africa; recipients of public funding act in a manner conducive to the public good; it acknowledges and rewards innovation; it enables economic growth through enterprise development; and it allows publishing of scientific results.[17] The challenge for public institutions like the ARC-SGI is that, traditionally, their research outputs have been in the public domain and the public sector had to ensure the widest dissemination of innovations.[5,15] Efforts to make public research institutions, such as the ARC-SGI, generate revenue by holding IPRs on plant varieties requires some exclusive licensing, such as is done by US Land Grant Universities.[5] According to Srinivasan[5], PVP/IPRs ownership by the public sector discourages exchanges of plant genetic resources and development of follow-on varieties. In addition, if exclusive rights are granted to larger companies, small businesses relying on public varieties (or varieties derived from public varieties) would be forced out of business. This would affect the development of the domestic wheat seed sector and create monopolies by big private companies. Concentration of ownership of plant variety rights in a few companies discourages follow-on innovations by other firms and researchers.[5]

The major role played by the ARC-SGI is supported by evidence from around the world indicating that plant-breeding innovations in non-hybrid crops like wheat have largely been sustained by the public sector. For example, Srinivasan[5] found that in cases where the public sector has played a major role in plant breeding, the overall concentration in ownership of plant variety rights was lower, particularly for non-hybrid crops which are less attractive to private sector investment. In addition, Eaton et al.[12] argue that although plant variety protection is expected to stimulate investments in plant breeding, for national agricultural research institutes (NARIs), PVP may address three objectives which are not always compatible: revenue collection, recognition of achievement and technology transfer. Their study found no evidence of potential revenue generation from plant breeding at national agricultural research institutes in Colombia and Uganda. In addition, one of the challenges for NARIs is to retain control of the plant breeding skills and resources for commercially important crops. For example, evidence from India and Kenya showed that the development of the private sector resulted in experienced NARI staff being hired into the private sector, making it difficult for the NARI to retain plant breeding staff and resources.

This implies that although the ARC-SGI is expected to generate revenue to sustain their activities, the concentration of potential revenue generation activities in more commercial crops brings into question public research resource allocations. Public research institutions would still be required to engage in research and plant breeding for mandated crops, such as staples, which might not be very lucrative compared with commercial crops. Such investments would have to be funded by public agricultural expenditures and other sources, making it difficult for NARIs such as the ARC-SGI to focus on revenue generation in their plant breeding activities. Also, the ARC-SGI faces competition from private companies, including in retaining its breeding skills and resources.

Conclusions and recommendations

This paper analysed the evolving landscape of wheat plant breeders' rights to address the dearth of empirical evidence on the patterns and trends of wheat varietal improvements in South Africa. The aim was to provide evidence on the evolution of varietal rights; the extent of varietal rights granted; changes of the rights on offer over time; changing ownership of the rights (including comparison between public and private, as well as domestic and foreign breeders) and; impacts of plant variety protection on wheat varietal development. The study compiled a detailed and novel count and attribute database of wheat varietal innovations in South Africa from 1979 to 2013, using information from the *South African Plant Variety Journal*, DAFF, South African National Library, and the ARC. The empirical analyses were based on descriptive statistics, trend analysis and graphical representation of trends and ownership of PBRs for wheat varietal improvements.

A total of 134 PBRs for wheat varietal innovations were lodged from 1979 to 2013, giving an average of 6 applications per year. This might have been driven by some breeders choosing not to apply for varietal protection, as well as by natural triggers such as outbreaks of pests and diseases, and different environmental and climatic conditions. The results indicate that plant breeders are increasingly seeking protection for their innovations. The number of PBR applications and grants for wheat variety innovations increased after the abolishment of the Wheat Board (6 applications per year compared with 4 per year before deregulation). The results also indicate that the administrative delays in granting PBR applications have been substantially reduced (by 77 days) post-deregulation, signifying increased efficiency in the processing of PBRs.

Since the publication of the *South African Plant Variety Journal* in 1979, the main applicants for wheat PVP have been Sensako (39%), ARC-SGI (25%) and Pannar (15%). After deregulation, Sensako's share decreased to 34%, while those of the ARC-SGI and Pannar increased by 5% and 9%, respectively. The results show that the ARC-SGI faces stiff competition from these well-established private companies. Establishing opportunities for collaboration with the private sector would encourage wheat variety innovation development. The advent of PVP/IPRs considerably restricted the exchange of plant genetic resources and breeding lines between the public and private research institutions. The ARC-SGI contributed to some of the PBRs owned by private companies through shared genetic resources made available before the PVP/IPRs were implemented. Future innovations and dissemination of wheat innovations could be stimulated by plant variety protection, together with the broader variety sector legislation that encourages both public and private sector investment.

Acknowledgements

This work was supported by the Bill and Melinda Gates Foundation grant to the University of Pretoria's Department of Agricultural Economics, Extension and Rural Development. The paper is part of the PhD research by Charity R. Nhemachena on: 'Biological innovations in South African agriculture: A study of wheat varietal change, 1950–2013' which was funded from the Bill and Melinda Gates Foundation grant.

Authors' contributions

C.R.N. conceptualised the article content and complied the first draft. F.G.L. and J.K. revised and made conceptual contributions, additions, and refined the article.

References

1. Kock MA, Gould C. 2013. Adapting IP to an evolving agricultural innovation landscape. Sniper No.2013/01028 : WIPO Magazine. 2013;2:25–29.
2. Claus P, Thiele-Witting M. 2002 Plant variety protection – a fascinating subject. World Patent Information. 2003;25:243–250. http://dx.doi.org/10.1016/S0172-2190(03)00074-7
3. Louwaars N, Dons H, Van Overwalle G, Raven H, Arundel A, Eaton D, et al. Breeding Business: The future of plant breeding in the light of developments in patent rights and plant breeders' rights. CGN Report No. 2009-14. Wageningen: Wageningen University, Centre for Genetic Resources; 2009.
4. Evenson RE, Gollin D. Crop variety improvement and its effect on productivity: The impact of international agricultural Research. Wallingford: CABI Publishing; 2003. http://dx.doi.org/10.1079/9780851995496.0000
5. Srinivasan CS. 2003. Concentration in ownership of plant variety rights: Some implications for developing countries. Food Policy. 2003;28:519–546. http://dx.doi.org/10.1016/j.foodpol.2003.10.003

6. Lesser W. Sector issue II: Seeds and plants. In: Siebeck EW, editor. Strengthening intellectual property rights in developing countries: A survey of literature, World Bank Discussion Paper No.112. Washington, DC: World Bank; 1990. p. 59–68.
7. Pardey PG, Koo B, Drew J, Horwich J, Nottenburg C. 2012. The evolving landscape of plant varietal rights in the United States, 1930–2008. Staff paper P12-1. St Paul, MN: University of Minnesota, College of Food, Agricultural and Natural Resource Sciences; 2012.
8. Pardey PG, Koo B, Drew J, Horwich J, Nottenburg C. 2013. The evolving landscape of plant varietal rights in the United States, 1930–2008. Nat Biotechnol. 2013;31(1):25–29. http://dx.doi.org/10.1038/nbt.2467
9. Stander CJ. The economics of cultivar improvement research in the South African wheat industry [MSc dissertation]. Pretoria: University of Pretoria; 2012.
10. International Union for the Protection of New Varieties of Plants (UPOV). International Convention for the Protection of New Varieties of Plants of 2 December 1961, as revised at Geneva on 10 November 1972, 23 October 1978 and 19 March 1991, Publication 221, Geneva: OPUV; 1994.
11. Tripp R, Louwaars N, Eaton D. 2007. Plant variety protection in developing countries. A report from the field. Food Policy. 2007;32:354–371. http://dx.doi.org/10.1016/j.foodpol.2006.09.003
12. Eaton D, Tripp R, Louwaars N. The effects of strengthened IPR regimes on the plant breeding sector in developing countries. International Agricultural Economist Conference; 2006 August 12–18; Gold Coast, Australia.
13. Alston JM, Venner RJ. The effects of the US Plant Variety Protection Act on wheat genetic improvement. Res Policy. 2002;31:527–542. http://dx.doi.org/10.1016/S0048-7333(01)00123-8
14. Srinivasan CS, Thirtle C. Potential economic impacts of terminator technologies: Policy implications for developing countries. Environ Dev Econ. 2003;8:187–205. http://dx.doi.org/10.1017/S1355770X0300010X
15. Diez MCF. The impact of plant varieties rights on research: The case of Spain. Food Policy. 2002;27:171–183. http://dx.doi.org/10.1016/S0306-9192(02)00010-6
16. Wright BD, Pardey PG. Changing intellectual property regimes: Implications for developing country agriculture. Int J Technol Glob. 2006;2(1/2):93–114. http://dx.doi.org/10.1504/IJTG.2006.009129
17. Moephuli SR, Moselakgomo MA, Phehane V. The Agricultural Research Council's role in plant variety protection and technology transfer – promoting publicly funded research. UPOV Workshop; 2012 March 05; Zanzibar, Tanzania.
18. Department of Agriculture, Forestry and Fisheries (DAFF). Plant Breeders' Rights policy, February 2011 [policy on the Internet]. c2011 [cited 2013 Jul 05]. Available from: http://www.nda.agric.za/docs/Policy/PlantBreederPol_2011.pdf.
19. RustTracker.org. A global wheat rust monitoring system [article on the Internet]. c2015 [cited 2015 Sep 01]. Available from: http://rusttracker.cimmyt.org.
20. Price SC. 2000.The public-private interface in plant breeding: Can there be a common culture? Diversity. 2000;15:6–7.

A bibliometric analysis of research on Ebola in Science Citation Index Expanded

AUTHORS:
Anastassios Pouris[1]
Yuh-Shan Ho[2]

AFFILIATIONS:
[1]Institute of Technological Innovation, University of Pretoria, Pretoria, South Africa
[2]Trend Research Centre, Asia University, Taichung, Taiwan

CORRESPONDENCE TO:
Yuh-Shan Ho

EMAIL:
ysho@asia.edu.tw

POSTAL ADDRESS:
Trend Research Centre, Asia University, 500 Lioufeng Road, Wufeng, Taichung 41354, Taiwan

DATES:
Received: 23 Aug. 2015
Revised: 14 Sep. 2015
Accepted: 14 Sep. 2015

KEYWORDS:
Ebola; Web of Science; Sci-Expanded; scientometrics; emerging virus

HOW TO CITE:
Pouris A, Ho YS, A bibliometric analysis of research on Ebola in Science Citation Index Expanded. S Afr J Sci. 2016;112(3/4), Art. #2015-0326, 6 pages. http://dx.doi.org/10.17159/sajs.2016/20150326

An unprecedented outbreak of the Ebola virus in 2014 claimed more than 1000 lives in West Africa and the World Health Organization declared a global public health emergency. This outbreak will undoubtedly promote additional research related to the Ebola virus and will create debate related to experimental drugs. This article identified the quantum of research in the field since 1991; the scientific disciplines that contributed to the field; the countries, organisations and authors that supported such research and the most cited articles. An increasing trend in annual production during 1991–2013 was observed. *Journal of Virology*, *Journal of Infectious Diseases*, and *Virology* were the three most productive journals in the field. Similarly, the field of virology dominated the 73 categories in which the Ebola research was classified. A total of 63 countries contributed to Ebola-related research, led by the USA. The most productive institutions were the United States Army Medical Research Institute of Infectious Diseases, the Centers for Disease Control and Prevention, and the National Institute of Allergy and Infectious Diseases. African countries were more likely to be involved in international collaboration than independent research. The most influential article exhibited a notable citation pattern and presented global trends in emerging infectious diseases.

Introduction

The worst Ebola outbreak in recorded history claimed more than 1000 victims[1] by 20 August 2014. It affected West Africa and Nigeria directly, but fears of its spreading affected the rest of the World. The World Health Organization called an international public health emergency and launched a USD 100 million response plan.[2]

The 2014 Ebola outbreak was first diagnosed in the remote southeastern area of Nzerekore in Guinea in February 2014, but subsequently spread to neighbouring Liberia, Sierra Leone and Nigeria.[3]

Ebola's origin is not known, but it is suspected that it is transmitted through forest bats. It can be transmitted between humans by touching sufferers or through body fluids.[4] The isolation of the aetiologic agent of Ebola haemorrhagic fever was reported for the first time during 1977.[5] According to the US Centers for Disease Control and Prevention,[4] since 1976 there have been 34 known outbreaks of Ebola. Before the latest outbreak, the deadliest episode was the first recorded outbreak in 1976 which killed 280 people in Central Africa.[4]

Apart from the mobilisation efforts to contain the virus and develop a relevant vaccine, the issue raised a number of ethical questions, as an experimental drug was used for two US health workers and not for African patients.[1]

The objective of this article is to report the results of an investigation aimed at mapping the research efforts related to Ebola virus internationally using scientometric techniques. These assessments are increasingly used as the basis for monitoring research performance of particular scientific disciplines and the support of appropriate policy actions. More specifically, the purpose of the study was to identify the state of Ebola virus research internationally and plot trends over time, to identify the institutions which undertake Ebola virus research and to pinpoint the scientific specialities which are emphasised in the field. The results of the investigation could possibly constitute benchmarks for monitoring the evolution of the research in this field and the impact of the outbreak on research efforts internationally.

Methodology

Following international best practice, we used evaluative scientometrics for the objectives of this investigation. Scientometrics is a tool by which the state of science and technology can be observed through the overall production of scientific literature at a given level of specialisation. Scientometric approaches have been used to map HIV/AIDS research[6,7], malaria research[8], and immunology[9].

Research articles indexed in the Science Citation Index Expanded (Sci-Expanded), a multidisciplinary database of the Thomson Reuters Web of Science, were identified for the objectives of this analysis (updated on 2014 August 05). The databases cover the most important international journals and the indices include the names and addresses of all co-authors, hence comprehensive coverage can be achieved. The keywords 'Ebola', 'Ebolas', and 'ebolavirus' were searched in the domain of topic (i.e. title, abstract, author keywords, and KeyWords Plus). KeyWords Plus provides search terms extracted from the titles of articles cited in Current Contents.[10] In order to ensure more exact information, another filter used was the 'front page'[11] approach. Only articles with the search keywords on their front page, including article title, abstract, and author keywords, were considered in this study and the time period examined was 1991–2013. The records were downloaded into a spreadsheet and were analysed using Microsoft Excel 2007. In the Sci-Expanded database, the corresponding author is designated as the 'reprint author'. In this study, the term 'corresponding author'[12] was used. The authors' corporate addresses determined the nationality of the article.

To investigate the citations received by the various articles, we estimated the number of times a paper was cited from Web of Science Core Collection. Total citations for a paper from its publication date to the end of a year, were recorded as $TC_{year.}$[13] Affiliations in England, Scotland, Northern Ireland, and Wales were reclassified as being from

the United Kingdom (UK). Affiliations in Hong Kong before 1997 and Macao before 1999 were included with China, while affiliations in USSR were amended to Russia.

Results and Discussion

Document type and language of publication

During the period under examination, 1623 documents from 12 document types were indexed. Of these documents, 70% were research articles, 10% reviews, 7.8% news items, 4.7% meeting abstracts, 4.5% editorial material, 3.9% proceedings papers, 1.6% letters, 0.62% corrections, 0.49% notes, 0.43% reprints, 0.12% book chapters and 0.062% biographical-items. The percentage of articles related to Ebola was lower than general medically-related fields, such as asthma in children[14] and Japanese lung cancer research[15], but was much higher than special infection research of severe acute respiratory syndrome[16] and *Helicobacter pylori*[17]. It was also noticeable that the document type, 'news items', took the third position. A high percentage of news items could also be found in severe acute respiratory syndrome (SARS) research[16], which was studied after its outbreak in 2003[18].

The 1133 articles were analysed more thoroughly and revealed that the majority (96%) were in English. The second most popular language was Russian with 2.2%, followed by French (11 articles), German (three articles) and Spanish (one article).

Figure 1 shows the number of relevant articles and citations per article for the period 1991–2013. The number of publications exhibited an increasing trend, albeit starting from a small base. During the recent years, there were approximately 100 articles produced per year. It is interesting to note that HIV/AIDS (another African disease) attracted more than 7000 articles per year.[7] It should however be emphasised that HIV/AIDS affects millions of people[7], while Ebola affected a few thousand patients for first time in the recorded history.

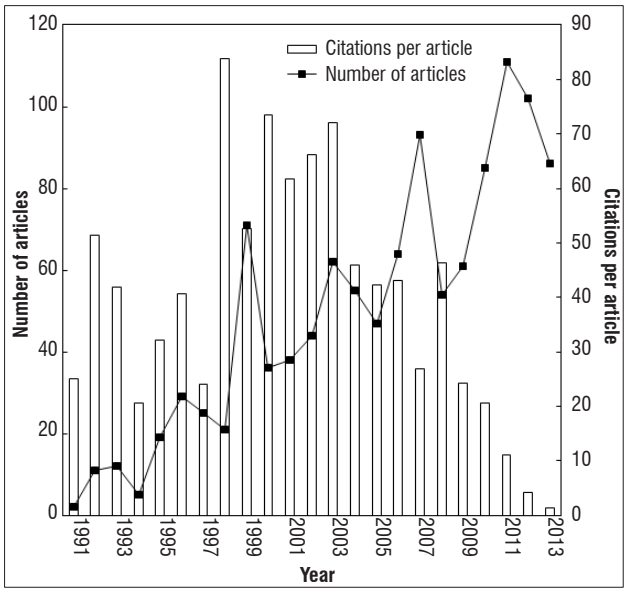


Figure 1: Ebola related research publications and citations per article 1991–2013.

Web of Science categories and journals

Distribution of Web of Science categories and journals have been studied in research topics.[19] In 2013, the Journal Citation Reports (JCR) indexed 8474 journals with citation references across 175 scientific disciplines in the science edition. The articles on Ebola were published in journals listed in 73 Web of Science categories in the science edition. During the period of the study, five categories published more than 100 articles, with the category virology contributing the most with 426 articles (38% of 1132 articles), followed by immunology (226 articles; 20%), microbiology (218; 19%), infectious diseases (203; 18%), and biochemistry and molecular biology (107; 9.5%). In 1999, 2007, and 2011, there were citation peaks in the categories immunology, microbiology and infectious diseases (Figure 2).

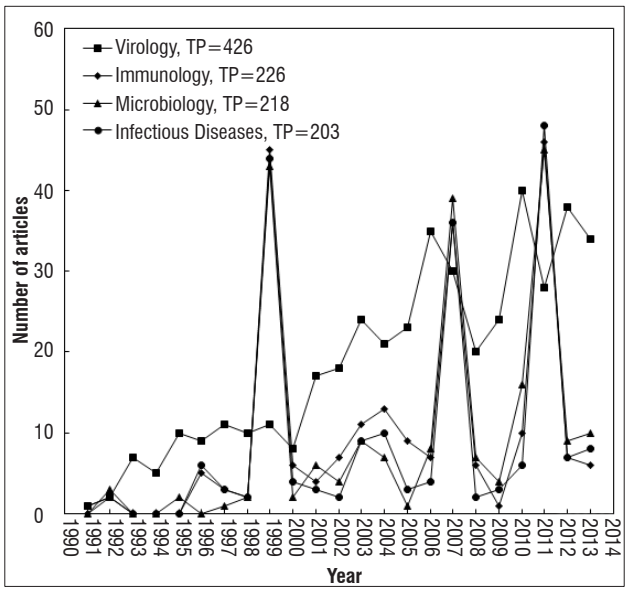


Figure 2: The number of articles in the five most productive Web of Science categories from 1991 to 2013.

The citation peak in 1999 could be because of the article 'Clinical virology of Ebola hemorrhagic fever (EHF): Virus, virus antigen, and IgG and IgM antibody findings among EHF patients in Kikwit, Democratic Republic of the Congo, 1995'[20] with TC_{2013}=119 in immunology and infectious diseases. The citation peak in 2007 could be attributed to the article 'The Ebola virus VP35 protein is a suppressor of RNA silencing'[21] with TC_{2013}=141 the category of microbiology.

The Ebola set of articles was further analysed according to the journals in which they were published. In total, 1133 articles were published in a range of 282 journals. According to Bradford's Law of Scattering[22], the journals were sorted in descending order in terms of number of articles, and then divided into three 'zones'. Zone 1 represented the most productive one-third of the total articles, with three (1.1%) of 282 journals. Zone 2 represented the next most productive one-third of total articles, with 28 (9.9%) of 282 journals (9.9%), and Zone 3 represented the least productive one-third of total articles with 251 (89%) of 282 journals. The three most productive core journals were *Journal of Virology,* which published the most proteome articles (172; 15% of 1133 articles), followed by *Journal of Infectious Diseases* (126; 11%), and *Virology* (71; 6.3%). In addition, regarding journal impact factor (IF), *Nature* gained first place with the highest IF_{2013} (42.351) with seven articles, followed by *Cell* (IF_{2013}=33.116) with two articles, *Science* (IF_{2013}=31.477) with eight articles, *Nature Medicine* (IF_{2013}=28.054) with 7 articles, and one article each for *Nature Immunology* (IF_{2013}=24.973) and *Nature Reviews Microbiology* (IF_{2013}=23.317).

Publications of country and institution

Excluding 11 articles without affiliation information of authors in Web of Science, there were 1122 articles with author information. Of these 1122 articles, 705 (63%) were single country articles and 417 (37%) were internationally collaborative articles. The top 10 countries, taking 93%, are listed in Table 1 under six indicators: total number of articles, single country articles, internationally collaborative articles, first author articles, corresponding author articles, and single author articles.[23] Of the articles ranked in the list of top 10 publications, 2 were from American countries, 5 were from European countries, 2 were Asian and 1 was from an African country, Gabon. Except for Italy, the Group of Eight (G8) (France, Germany, Italy, Japan, the UK, the USA, Canada, and Russia) were ranked in the top 7. This is no coincidence, as the earliest Ebola research took place in these industrialised countries

(e.g. 'A case of Ebola virus-infection'[24] and 'Ebola virus: Comparison, at ultrastructural level, of the behaviour of the Sudan and Zaire strains in monkeys').[25]

Table 1: Ebola virus research: Country of publication

Country	TP	TPR (%)	IPR (%)	CPR (%)	FPR (%)	RPR (%)	SPR (%)
USA	772	1 (68)	1 (67)	1 (71)	1 (57)	1 (57)	1 (53)
Germany	154	2 (14)	3 (4.9)	2 (29)	2 (6.9)	2 (6.8)	4 (3.9)
Canada	112	3 (9.9)	4 (3.4)	3 (21)	5 (4.2)	5 (4.1)	N/A
France	104	4 (9.2)	5 (3.0)	5 (20)	3 (4.8)	3 (4.9)	4 (3.9)
Japan	96	5 (8.5)	9 (1.1)	3 (21)	4 (4.7)	4 (4.6)	8 (2.0)
UK	60	6 (5.3)	6 (2.3)	6 (11)	7 (2.3)	7 (2.2)	2 (7.8)
Russia	53	7 (4.7)	2 (5.8)	18 (2.9)	6 (3.8)	6 (3.8)	N/A
Gabon	42	8 (3.7)	15 (0.42)	7 (9.4)	7 (2.3)	7 (2.2)	N/A
Belgium	40	9 (3.5)	10 (1.0)	8 (7.9)	9 (1.8)	9 (1.8)	4 (3.9)
Spain	29	10 (2.6)	7 (1.8)	14 (3.8)	10 (1.5)	10 (1.6)	8 (2.0)

TP, total articles; TPR (%), rank and the percentage of total articles; IPR (%), rank and the percentage of independent articles; CPR (%), rank and the percentage of internationally collaborative articles; FPR (%), rank and the percentage of first author articles; RPR (%), rank and the percentage of the corresponding authored articles; SPR (%), rank and the percentage of single author articles; N/A, not available.

The USA dominated the table producing 68% of the articles during the period. It ranked top in all six indicators and was followed distantly by other countries. The USA had the most-frequent partners, accounting for 71% of all the internationally collaborative articles. Gabon was ranked in 8th position but had no single author articles. A number of African countries appeared to produce Ebola related publications including Uganda (25 articles), Congo (21), Zaire (18), South Africa (18), Central African Republic (7), Cote Ivoire (7), Cameroon (6), Zambia (6), Kenya (4), Ghana (3), Sierra Leone (2), Tanzania (2), Zimbabwe (2), and 1 article each from Angola, Burkina Faso, Egypt, Guinea, Rwanda, Senegal and Sudan.

The contribution of different institutions was estimated by their affiliation to at least 1 author of the Ebola articles. Of the 1122 Ebola articles analysed, 327 (29% of 1122 articles) were single institution publications and 795 (71%) were inter-institutionally collaborative publications. The top 10 institutions were ranked according to their number of articles. This included the number of single institution articles and inter-institutionally collaborative articles; first author articles; and single author articles (Table 2). Out of the top 11 most productive institutions, seven were affiliated to the USA, two were from Canada, and one each from Germany and Japan, respectively. The United States Army Medical Research Institute of Infectious Diseases in the USA was the most productive institution, followed by Centers for Disease Control and Prevention and National Institute of Allergy and Infectious Diseases (NIAID) in the USA. Centers for Disease Control and Prevention and National Institute of Allergy published the most single author articles. The top 10 productive institutions in Africa were Centre International de Recherches Médicales de Franceville in Gabon (32 articles), Uganda Virus Research Institute in Uganda (13), Ministry of Health in Uganda (8), Ministry of Health in the Democratic Republic of Congo (6), National Institute for Virology in South Africa (6), National Institute for Communicable Diseases in South Africa (6), University of Zambia in Zambia (6), Institut de Recherche pour le Développement in Gabon (5), Kikwit General Hospital in Congo (5), and Institut Pasteur in Central African Republic (5).

Table 2: Prolific institutions in Ebola research

Institution	TP	CPR (%)	FPR (%)	SPR (%)
United States Army Medical Research Institute of Infectious Diseases, USA	181	1 (11)	1 (7.4)	6 (2.0)
Centers for Disease Control and Prevention, USA	126	3 (6.3)	3 (5.0)	1 (7.8)
National Institute of Allergy and Infectious Diseases (NIAID), USA	106	5 (3.6)	2 (5.2)	3 (3.9)
University of Marburg, Germany	85	4 (4.2)	5 (2.4)	N/A
University of Manitoba, Canada	73	N/A	55 (0.27)	N/A
Public Health Agency of Canada, Canada	72	18 (0.60)	6 (2.1)	N/A
University of Tokyo, Japan	68	44 (0.30)	7 (2.0)	N/A
University of Wisconsin, USA	56	N/A	9 (1.8)	N/A
University of Pennsylvania, USA	51	2 (7.2)	4 (3.3)	6 (2.0)
Harvard University, USA	33	44 (0.30)	19 (0.71)	6 (2.0)
Scripps Research Institute, USA	33	7 (1.8)	9 (1.8)	6 (2.0)

TP, total articles; CPR (%), rank and the percentage of inter-institutionally collaborative articles; FPR (%), rank and the percentage of first author articles; SPR (%), rank and the percentage of single author articles; N/A not available.

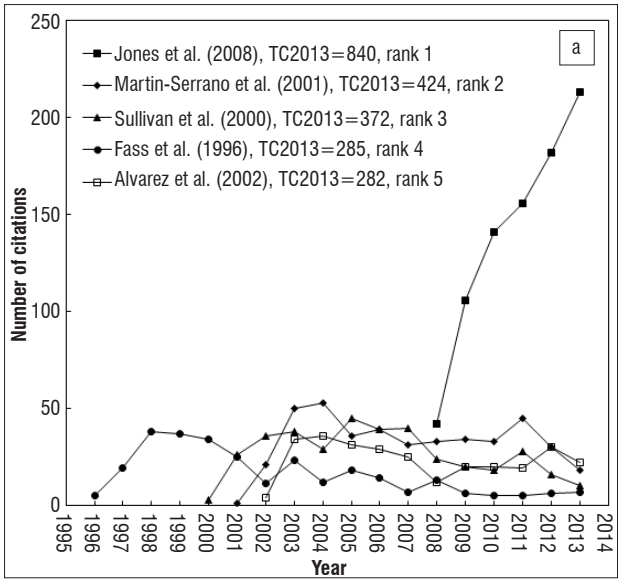
Highly cited articles

The total number of times an article was cited from Web of Science Core Collection since its publication to the end of 2013 (TC_{2013}) was estimated.[13] TC_{2013}, an accumulative number, could reach a large value as long as the time span is long enough. The citation lives of the top 10 articles ($TC_{2013}>250$) are shown in Figure 3. The only 1 article that kept a sharply increasing citation trend after its publication was 'Global trends in emerging infectious diseases'[26] published in by 7 authors from Wildlife Trust in China, Zoological Society of London in the UK, Columbia University in the USA, and University of Georgia in the USA. In this article, 335 emerging infectious diseases (EID) events were analysed and the authors demonstrated non-random global patterns between 1940 and 2004. A conclusion was that EID events are dominated by zoonoses (60.3% of EIDs), with the majority of these (71.8%) originating in wildlife (for example, severe acute respiratory virus, Ebola virus), and increasing significantly over time. Global trends in emerging infectious diseases'[26] not only discussed Ebola but also other emerging infectious diseases. Table 3 presents the top 21 articles cited more than 200 times ($TC_{2013}>200$). Out of these 21 articles, eight (38% of 21 articles) were published in the 1990s and 13 (62%) in the 2000s. The first article, which was cited more than 200 times, was published in 1996. The latest one was published in 2008. The journals in which these articles were published were *Proceedings of the National Academy of Sciences of the United States of America* ($IF_{2013}=9.809$) with six articles, followed by *Nature* (four articles; $IF_{2013}=42.351$), *Nature Medicine* (two articles; $IF_{2013}=28.054$), *Journal of Virology* (two articles; $IF_{2013}=4.648$), and one for each of *Bulletin of the World Health Organization* ($IF_{2013}=5.112$), *Journal of Clinical Microbiology* ($IF_{2013}=4.232$), *Journal of Experimental Medicine* ($IF_{2013}=13.912$), *Molecular Cell* ($IF_{2013}=14.464$), *Molecular Membrane Biology* ($IF_{2013}=1.729$), *Nature Structural Biology* ($IF_{2013}=11.579$), and *Science* ($IF_{2013}=31.477$), respectively. *Nature Structural Biology* was not listed in Sci-Expanded after 2003. In 1998 and 1999, Weissenhorn from European Molecular Biology Laboratory in France published 2 highly cited papers as first author: 'Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain'[27] with $TC_{2013}=254$ and 'Structural basis for membrane fusion by enveloped viruses'[28] with $TC_{2013}=256$. In 2000 and 2003, both Sullivan from National Institutes of Health in the USA and Basler from Mount Sinai School of Medicine in the USA also published 2 highly cited papers as first authors.

Table 3: Top 21 most cited articles (TC_{2013}>200)

Rank	TC_{2013}	Reference
1	840	Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990–993.
2	424	Martin-Serrano J, Zang T, Bieniasz PD, HIV-I and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. Nat Med. 2001;7(12):1313–1319.
3	372	Sullivan NJ, Sanchez A, Rollin PE, Yang Z, Nabel GJ. Development of a preventive vaccine for Ebola virus infection in primates. Nature. 2000; 408(6812):605–609.
4	285	Fass D, Harrison, SC, Kim PS. Retrovirus envelope domain at 1.7 angstrom resolution. Nat Struct Biol. 1996;3(5):465–469.
5	282	Alvarez CP, Lasala F, Carrillo J, Muniz O, Corbi AL, Delgado R. C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans. J Virol. 2002;76(13):6841–6844.
6	275	Simonsen L, Kane A, Lloyd J, Zaffran M, Kane M. Unsafe injections in the developing world and transmission of bloodborne pathogens: A review. B World Health Organ. 1999;77(10):789–800.
7	262	Chandran K, Sullivan NJ, Felbor U, Whelan SP, Cunningham JM. Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. Science, 2005;308(5728):1643–1645.
8	261	Takada A, Robison C, Goto H, Sanchez A, Murti KG, Whitt MA. Kawaoka YA system for functional analysis of Ebola virus glycoprotein. P Nat Acad Sci USA. 1997;94(26):14764–14769.
9	256	Weissenhorn W, Dessen A, Calder LJ, Harrison SC, Skehel JJ, Wiley DC. Structural basis for membrane fusion by enveloped viruses. Mol Memb Biol. 1999;16(1):3–9.
10	254	Weissenhorn W, Carfi A, Lee KH, Skehel JJ, Wiley DC. Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. Mol Cell. 1998;2(5):605–616.
11	248	Bavari S, Bosio CM, Wiegand E, Ruthel G, Will AB, Geisbert TW, et al. Lipid raft microdomains: A gateway for compartmentalized trafficking of Ebola and Marburg viruses. J Exp Med. 2002;195(5):593–602.
12	244	Sanchez A, Trappier SG, Mahy BWJ, Peters CJ, Nichol ST. The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. P Nat Acad Sci USA. 1996;93(8):3602–3607.
13	241	Sullivan NJ, Geisbert W, Geisbert JB, Xu L, Yang ZY, Roederer, et al. Accelerated vaccination for Ebola virus haemorrhagic fever in non-human primates. Nature. 2003;424(6949):681–684.
14	240	Basler CF, Wang XY, Muhlberger E, Volchkov V, Paragas J, Klenk HD. The Ebola virus VP35 protein functions as a type IIFN antagonist. P Nat Acad Sci USA. 2000;97(22):12289–12294.
15	233	Harty RN, Brown ME, Wang G, Huibregtse J, Hayes FP. A PPxY motif within the VP40 protein of Ebola virus interacts physically and functionally with a ubiquitin ligase: Implications for filovirus budding. P Nat Acad Sci USA. 2000;97(25):13871–13876.
16	219	Baize S, Leroy EM, Georges-Courbot MC, Capron M, Lansoud-Soukate J, Debre P. Defective humoral responses and extensive intravascular apoptosis are associated with fatal outcome in Ebola virus-infected patients. Nat Med. 1999;5(4):423–426.
17	214	Strack B, Calistri A, Accola MA, Palu G, Gottlinger HG. A role for ubiquitin ligase recruitment in retrovirus release. P Nat Acad Sci USA. 2000;97(24):13063–13068.
18	212	Drosten C, Gottig S, Schilling S, Asper M, Panning M, Schmitz H, et al. Rapid detection and quantification of RNA of Ebola and Marburg viruses, Lassa virus, Crimean-Congo hemorrhagic fever virus, Rift Valley fever virus, Dengue virus, and Yellow fever virus by real-time reverse transcription-PCR. J Clin Microbiol. 2002;40(7):2323–2330.
18	212	Volchkov VE, Feldmann H, Volchkova VA, Klenk HD. Processing of the Ebola virus glycoprotein by the proprotein convertase furin. P Nat Acad Sci USA. 1998;95(10):5762–5767.
20	210	Walsh PD, Abernethy KA, Bermejo M, Beyersk R, De Wachter P, Akou ME, et al. Catastrophic ape decline in western equatorial Africa. Nature. 2003;422(6932):611–614.
21	209	Basler CF, Mikulasova A, Martinez-Sobrido L, Paragas J, Muhlberger E, Bray M, et al. The Ebola virus VP35 protein inhibits activation of interferon regulatory factor 3. J Virol. 2003;77(14):7945–7956.

TC_{2013}, total number of times the article was cited from Web of Science Core Collection since its publication to the end of 2013.



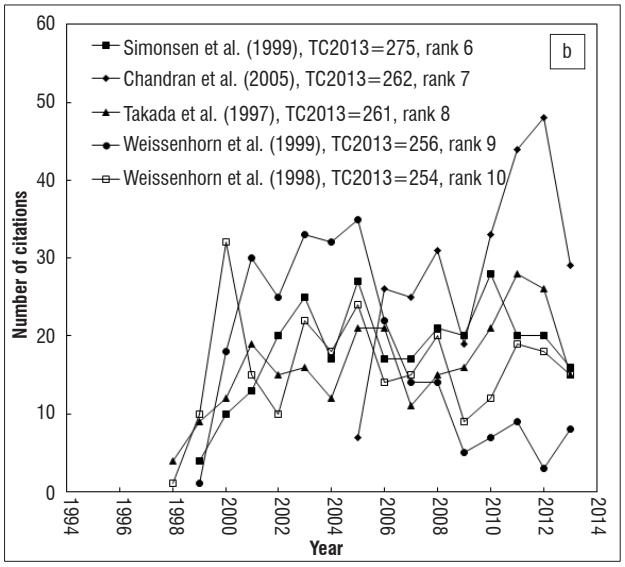


Figure 3: (a) The life of the top 5 most cited articles (TC_{2013}>280) and (b) the top 6–10 most cited articles (280 > TC_{2013} > 250).

Conclusions

This article provided for the first time an overview of research related to Ebola virus undertaken internationally and identified that there has been limited international research. The Ebola virus-related articles showed an increasing trend during 1991–2013. Virology was the most dominant discipline in the domain, and *Journal of Virology*, *Journal of Infectious Diseases*, and *Virology* were the most favoured journals. Industrialised countries contributed to more than 70% of the articles, while a number of African countries played an important role in international collaboration. The United States Army Medical Research Institute of Infectious Diseases, the Centers for Disease Control and Prevention, and the National Institute of Allergy and Infectious Diseases were the most prolific institutions researching the topic. The top 3 most cited articles in the field were published in *Nature*. The most cited article with a distinguished citation pattern included hundreds of emerging infectious diseases events and non-random global patterns from 1940–2004.

The Ebola crisis has raised a number of issues that are subject to further research. How can the authorities decide on the resources to be spent on various diseases and related prevention activities? A controversy about drug testing in Africa and who should benefit for limited experimental drugs are also complexities facing public health authorities internationally outside the critical issue of how to control the disease and prevent its spreading to other countries. Similarly, it will be interesting to identify how the outbreak will affect research activities in the field in the foreseeable future.

Authors' contributions

The authors made equal contributions to this article.

References

1. Cowell A. Ebola death toll in West Africa tops 1200. The New York Times. 2014 Aug 19. Available from: http://www.nytimes.com/2014/08/20/world/africa/ebola-outbreak.html?_r=0.2014.
2. World. Health Organization (WHO). Statement on the Meeting of the International Health Regulations Emergency Committee regarding the 2014 Ebola outbreak in West Africa. Geneva: WHO; 2014. Available from: http://www.who.int/mediacentre/news/statements/2014/ebola-20140808/en/
3. Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, et al. Emergence of Zaire Ebola virus disease in Guinea — Preliminary report. N Engl J Med. 2014;371:1418–1425. http://dx.doi.org/10.1056/NEJMoa1404505
4. Centers for Disease Control and Prevention (CDC). Outbreaks Chronology: Ebola Virus Disease. Atlanta, GA: CDC; 2014. Available from: http://www.cdc.gov/vhf/ebola/resources/outbreak-table.html.
5. Bowen ETW, Platt GS, Simpson DIH, Mcardell LB, Raymond RT. Ebola hemorrhagic fever: Experimental infection of monkeys. T Roy Soc Trop Med H. 1978;72(2):188–191. http://dx.doi.org/10.1016/0035-9203(78)90058-5
6. Small H. A SCI-map case-study: Building a map of AIDS research. Scientometrics. 1994;30(1):229–241. http://dx.doi.org/10.1007/BF02017225
7. Pouris A, Pouris A. Scientometrics of a pandemic: HIV/AIDS research in South Africa and the World. Scientometrics. 2011;86(2):541–552. http://dx.doi.org/10.1007/s11192-010-0277-6
8. Garg KC, Kumar S, Madhavi Y, Bahl M. Bibliometrics of global malaria vaccine research. Health Info Libr J. 2009;26(1):22–31. http://dx.doi.org/10.1111/j.1471-1842.2008.00779.x
9. Dos Santos NF, Rumjanek VM. Brazilian immunology: One hundred years later. Scientometrics. 2001;50(3):405–418. http://dx.doi.org/10.1023/A:1010502513971
10. Garfield E. KeyWords PlusTM - ISIS breakthrough retrieval method. 1. Expanding your searching power on current-contents on diskette. Curr Comments. 1990;32:5–9.
11. Fu HZ, Wang MH, Ho YS. The most frequently cited adsorption research articles in the science citation index (expanded). J Colloid Interf Sci. 2012;379(1):148–156. http://dx.doi.org/10.1016/j.jcis.2012.04.051
12. Ho YS. Top-cited papers in chemical engineering in Science Citation Index Expanded: A bibliometric analysis. Chinese J Chem Eng. 2012;20(3):478–488. http://dx.doi.org/10.1016/S1004-9541(11)60209-7
13. Chuang KY, Wang MH, Ho YS. High-impact papers presented in the subject category of water resources in the essential science indicators database of the Institute for Scientific Information. Scientometrics. 2011;87(3):551–562. http://dx.doi.org/10.1007/s11192-011-0365-2
14. Chen SR, Chiu WT, Ho YS. Asthma in children: Mapping the literature by bibliometric analysis. Rev Fr Allerg Immun Clin. 2005;45(6):442–446. http://dx.doi.org/10.1016/j.allerg.2005.08.002
15. Ho YS, Satoh H, Lin SY. Japanese lung cancer research trends and performance in science citation index. Internal Med. 2010;49(20):2219–2228. http://dx.doi.org/10.2169/internalmedicine.49.3687
16. Chiu WT, Huang JS, Ho YS. Bibliometric analysis of severe acute respiratory syndrome-related research in the beginning stage. Scientometrics. 2004;61(1):69–77. http://dx.doi.org/10.1023/B:SCIE.0000037363.49623.28
17. Suk FM, Lien GS, Yu TC, Ho YS. Global trends in Helicobacter pylori research from 1991 to 2008 analyzed with the Science Citation Index Expanded. Eur J Gastroen Hepat. 2011;23(4):295–301. http://dx.doi.org/10.1097/MEG.0b013e3283457af7
18. World Health Organization (WHO). Cumulative number of reported probable cases of SARS from 1 Nov 2002 to 6 June 2003 [report on the Internet]. c2016 [cited 2003 Jul 19]. Available from: http://www.who.int/csr/sars/country/2003_06_06/en/

19. Chiu WT, Ho YS. Bibliometric analysis of homeopathy research during the period of 1991 to 2003. Scientometrics. 2005;63(1):3–23. http://dx.doi.org/10.1007/s11192-005-0201-7

20. Ksiazek TG, Rollin PE, Williams AJ, Bressler DS, Martin ML, Swanepoel R, et al. Clinical virology of Ebola hemorrhagic fever (EHF): Virus, virus antigen, and IgG and IgM antibody findings among EHF patients in Kikwit, Democratic Republic of the Congo, 1995. J Infect Dis. 1999;179(S1):S177–S187. http://dx.doi.org/10.1086/514321

21. Haasnoot J, De Vries W, Geutjes EJ, Prins M, De Haan P, Berkhout B. The Ebola virus VP35 protein is a suppressor of RNA silencing. PLoS Pathog. 2007;3(6):e86. http://dx.doi.org/10.1371/journal.ppat.0030086

22. Bradford SC. Sources of information on specific subjects. Brit J Eng. 1934;137(3550):85–86.

23. Ho YS, Kahn M. A bibliometric study of highly cited reviews in the Science Citation Index Expanded™. J Assoc Inform Sci Technol. 2014;65(2):372–385. http://dx.doi.org/10.1002/asi.22974

24. Emond RTD, Evans B, Bowen ETW, Lloyd G. A case of Ebola virus-infection. Brit Med J. 1977;2(6086):541–544. http://dx.doi.org/10.1136/bmj.2.6086.541

25. Ellis DS, Bowen ETW, Simpson DIH, Stamford S. Ebola virus: Comparison, at ultrastructural level, of the behavior of the Sudan and Zaire strains in monkeys. Brit J Exp Pathol. 1978;59(6):584–593.

26. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990–993. http://dx.doi.org/10.1038/nature06536

27. Weissenhorn W, Carfi A, Lee KH, Skehel JJ, Wiley DC. Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. Mol Cell. 1998;2(5):605–616. http://dx.doi.org/10.1016/S1097-2765(00)80159-8

28. Weissenhorn W, Dessen A, Calder LJ, Harrison SC, Skehel JJ, Wiley DC. Structural basis for membrane fusion by enveloped viruses. Mol Membr Biol. 1999;16(1):3–9. http://dx.doi.org/10.1080/096876899294706

Recent trends in the climate of Namaqualand, a megadiverse arid region of South Africa

AUTHORS:
Claire L. Davis[1]
M. Timm Hoffman[2]
Wesley Roberts[3]

AFFILIATIONS:
[1]Natural Resources and the Environment, Council for Scientific and Industrial Research, Stellenbosch, South Africa
[2]Department of Biological Sciences, University of Cape Town, Cape Town, South Africa
[3]BioCarbon Partners, Lusaka, Zambia

CORRESPONDENCE TO:
Claire Davis

EMAIL:
cdavis@csir.co.za

POSTAL ADDRESS:
Natural Resources and the Environment, Council for Scientific and Industrial Research, PO Box 320, Stellenbosch 7599, South Africa

DATES:
Received: 05 June 2015
Revised: 05 Aug. 2015
Accepted: 23 Aug. 2015

KEYWORDS:
climate extreme; trend analysis; temperature; rainfall; evapotranspiration

HOW TO CITE:
Davis CL, Hoffman MT, Roberts W. Recent trends in the climate of Namaqualand, a megadiverse arid region of South Africa. S Afr J Sci. 2016;112(3/4), Art. 2015-0217, 9 pages. http://dx.doi.org/10.17159/sajs.2016/20150217

Namaqualand is especially vulnerable to future climate change impacts. Using a high-resolution (0.5°x0.5°) gridded data set (CRU TS 3.1) and individual weather station data, we demonstrated that temperatures as well as frequency of hot extremes have increased across this region. Specifically, minimum temperatures have increased by 1.4 °C and maximum temperatures by 1.1 °C over the last century. Of the five weather stations analysed, two showed evidence of a significant increase in the duration of warm spells of up to 5 days per decade and a reduction in the number of cool days (TX10P) by up to 3 days per decade. In terms of rainfall, we found no clear evidence for a significant change in annual totals or the frequency or intensity of rainfall events. Seasonal trends in rainfall did, however, demonstrate some spatial variability across the region. Spatial trends in evapotranspiration obtained from the 8-day MOD16 ET product were characterised by a steepening inland-coastal gradient where areas along the coastline showed a significant increase in evapotranspiration of up to 30 mm per decade, most notably in spring and summer. The increase in temperature linked with the increases in evapotranspiration pose significant challenges for water availability in the region, but further research into changes in coastal fog is required in order for a more reliable assessment to be made. Overall, the results presented in this study provide evidence-based information for the management of climate change impacts as well as the development of appropriate adaptation responses at a local scale.

Introduction

There is strong scientific evidence that recent changes in climate are probably attributable to human activities and have resulted in increased annual global temperatures, as well as associated increases in temperature extremes.[1,2] There have been a number of studies across South Africa that have demonstrated similar trends in temperature.[1,3-9] Where records are of sufficient length, there have been detectable increases in extreme rainfall events[1,4,9] and evidence that droughts have become more intense and widespread.[6] The latest climate change projections for South Africa indicate that temperatures are likely to increase into the 21st century and that rainfall in the south western Cape is expected to decrease in the future as a result of a poleward retreat of rain-bearing mid latitude cyclones.[9] These projected changes in climate are likely to have significant implications for the region's biodiversity.[10]

The Succulent Karoo biome, including the Namaqualand region, is globally recognised as a biodiversity hotpot that is particularly vulnerable to climate change,[11] not only because of the high levels of biodiversity but also because of a high dependence of the region's population on natural resources, livestock production, and dryland agriculture.[12,13] Modelling studies have indicated that the Succulent Karoo may suffer a reduction in spatial extent of up to 40% as well as consequent reductions in the abundance and diversity of endemic species as a result of changes in temperature and rainfall.[14,15] In terms of individual species' responses to climate change, a study by Broennimann et al.[16] found that geophytes and succulents, which make up over half of plant species in Namaqualand, were particularly vulnerable to climate change. They predicted a minimum decline in species richness of 41% for the Succulent Karoo biome.

To date, only three studies have investigated the historical climate of the Namaqualand,[17-19] all of which focused specifically on changes in rainfall. Various palaeoclimate studies in the winter rainfall zone of South Africa[20-23] have also contributed to the understanding of the long-term changes in climate in Namaqualand. Distinguishing local climate trends is essential as climate may not change uniformly across large areas. Furthermore, there is a clear demand for reliable climate information by both local authorities and land managers in order to ensure that climate risk management and assessments of climate change are locally relevant.[24,25]

We present an assessment of the trends in key climatic variables as well as climate extremes for Namaqualand that provide an updated analysis, add value and complement previous studies. A better understanding of local changes and inter-annual variability assist in the interpretation of future climate change projections and provide evidence-based research to guide management actions in the region.

Detecting changes in climate at the local scale is considerably more difficult than doing so for regional climate because of the lack of an accurate, long-term, well-maintained and dense spatial network of observational stations to detect regional climate signals. This is particularly evident in Namaqualand where weather stations are sparsely distributed. Gridded climate data sets developed by interpolating weather station records help overcome this by approximating the true spatial and temporal variability of key climate variables.[26-29] The analysis presented in this paper is based on temperature and rainfall data from the high-resolution gridded data set (1901–2009) provided by the Climatic Research Unit of the University of East Anglia (CRU TS 3.1) as well as daily climate data obtained from five weather stations in Namaqualand and evapotranspiration data obtained from the 8 day MOD16 product. Multiple data sets were used in order to explore the full range of climate signals and to provide a complete characterisation of the confidence in the changes observed.

Study Area

Namaqualand is located in the north-west corner of South Africa and extends from the Orange River in the north to Vanrhynsdorp and the Olifants River in the south.[30,31] It covers a region of approximately 50 000 km^2 and is located west and south of the escarpment, 200–300 km inland of the west coast[32] (Figure 1).

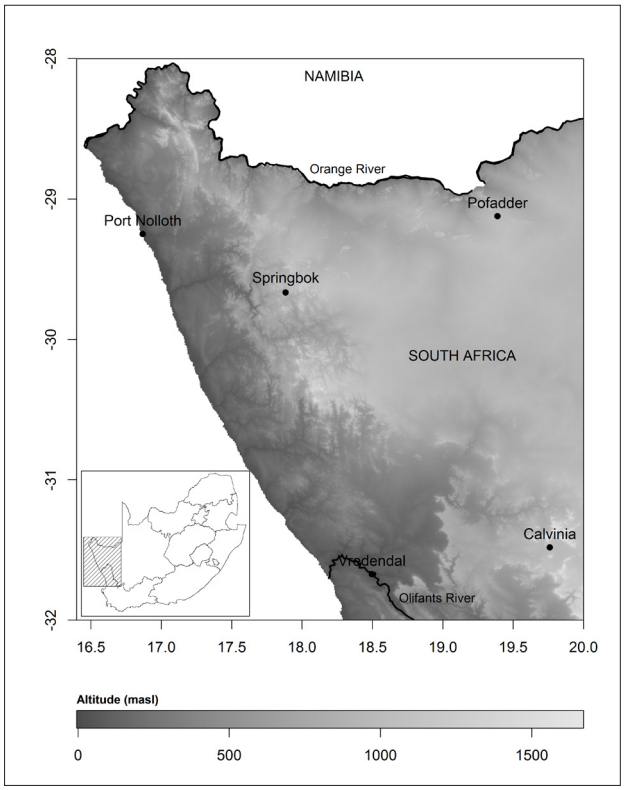


Figure 1: Map of Namaqualand showing altitude (Digital Elevation Datatbase (DEM), SRTM version 4, 250 m) and five weather stations used in the analysis of trends in daily climate extremes. For this study, Namaqualand has been defined as the area between the Orange River in the north and the Olifants River in the South.[31,37]

Namaqualand hosts approximately 3500 plant species with about 25% of this flora being endemic to the region[31] and is well known for its flower displays in spring. Agriculture, mainly livestock production, is the primary land use in Namaqualand and is practised on both commercial and communal farmlands.[33] The communal areas account for more than 25% of the region and support 45% of its population.[34]

The climate of the region is determined primarily by the southern subtropical high pressure system and the circumpolar westerly airstream.[35] Geographic features, such as the mountains of the escarpment and the cold Benguela current, influence local scale climate. Namaqualand is classed as a semi-arid winter rainfall region with mean annual precipitation ranging from 50 mm in the north-west to 400 mm per annum in the Kamiesberg. Peak rainfall amounts occur over the Kamiesberg Mountains as a result of the orographic effect.[17] More than 60% of the rainfall occurs during the winter months between May and September as a result of the cold, westerly fronts from the southern oceans.[35] Coastal lows are also common in winter. Rainfall is supplemented by heavy dewfalls experienced during mid-winter (July–August) and advective coastal fog experienced during the summer months. Fog is generated by the cold Benguela current of the Atlantic Ocean and occurs primarily along the coastal region for about 75 days of the year.[36] Namaqualand is prone to droughts, which usually span a few successive years.[17,19]

Annual average temperatures for the study area are relatively mild throughout the year and range from 13 °C to 21 °C, owing to the cold Benguela Current off the west coast of Namaqualand.[31] Mean annual temperatures are highest inland of the west coast increasing northwards, while the escarpment and high lying areas experience cooler temperatures. Maximum temperatures only exceed 30 °C when Berg winds[35] are blowing off the plateau to the west. There is a relatively high evaporative demand in Namaqualand, particularly during these hot berg-wind events.[36] With the exception of the coastal belt, the region experiences a large annual and diurnal range in temperatures.[17] Occasional frosts occur in the high lying areas of the escarpment and central plateau.[32]

Methodology

Observed changes in climate

A high-resolution (0.5°x0.5°) gridded monthly data set provided by the CRU TS 3.1[29] was used to analyse the spatial trends in temperature and rainfall for the period 1901–2009. Acknowledging the limitations of weather station interpolation, a gridded climate data set was used in this study as the weather stations are sparsely distributed in Namaqualand and are thus not fully representative of the full range of temperature and rainfall values experienced over the region. The CRU TS 3.1 data set has been validated,[28] and was used previously to study trends in African climate.[5] It is important to note that gridded climate data is not suitable to provide time series analysis at the scale of the individual pixel and thus the results presented here provide an indication of the spatial patterns on change in temperature and rainfall over the entire region. A time series of temperature and rainfall data was extracted from the CRU TS 3.1 data set and a linear regression was performed to detect changes at an annual time step for the region. To describe inter-annual variability, temperature and rainfall anomalies were computed by subtracting the 1960–1990 mean value[38,39] from each annual value in the time series.

In order to ensure that the CRU TS 3.1 data set depicts the correct distribution and amounts of monthly temperature and rainfall for the studied region, the gridded data set was compared to the weather station data (Table 1) for the period 1982–2009 by calculating the correlation coefficients (r-value). With the exception of minimum temperature obtained from the Springbok and Port Nolloth stations, the weather station data and the gridded CRU TS 3.1 were significantly and positivity correlated (Supplementary Table 1). The correlation coefficients (>0.5) confirm the similarity of the two data sets, thus lending confidence in using the interpolated data set for the spatial analysis of trends.

Table 1: List of weather stations utilised in the analysis

Station number	Station name	Latitude (°S)	Longitude (°E)	Altitude (m)	Period
0134479A3	Calvinia	-31.4820	19.7610	977	1986–2014
0247668A4	Pofadder	-29.1230	19.3890	984	1982–2014
02426446	Port Nolloth	-29.2500	16.8680	8	1983–2014
0214670A0	Springbok	-29.6670	17.8830	1006	1987–2013
0106880A2	Vredendal	-31.6730	18.4960	42	1982–2014

In order to provide a more detailed understanding of changes in water availability for Namaqualand, trends in evapotranspiration were also assessed. A direct measurement of evapotranspiration was not carried out in this study, but was derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) on-board NASA's Terra and Aqua satellites. The 8-day MOD16 Global Terrestrial Evapotranspiration data set (MOD 16 ET available from http://www.ntsg.umt.edu/project/mod16), has a 1 km^2 spatial resolution and covers the time period 2000–2010. The MOD16 Evapotranspiration data sets are estimated using the improved evapotranspiration algorithm outlined in Mu et al.[40] which is based on the Penman-Monteith equation.[41]

Trend Analysis

Spatial trends in temperature derived from the CRU TS 3.1 data set (1901–2009) and evapotranspiration derived from the MODIS MOD16 ET (2000–2010) data set were examined by applying the Theil–Sen (TS) median slope estimator.[42,43] This is a robust non-parametric trend operator which is unaffected by the presence of outliers.[44] The Theil-Sen operator is determined by calculating the slope between every pair-wise combination and then assessing the median slope value. The result is a spatially-explicit expression of the rate of change per year in each of the climate variables, where pixels with negative (positive) slopes indicate areas that have experienced a decline (increase). The trend analysis was calculated for seasonal means. The four major seasons are December-January-February (DJF), March-April-May (MAM), June-July-August (JJA), and September-October-November (SON). For comparison, trends in annual temperature and rainfall for each of the weather stations were computed.

To assess changes in rainfall the rank-based, non-parametric Mann-Kendall method[45] was applied. This is a non-linear trend indicator that measures the degree to which a trend is consistently increasing or decreasing. The Kendall's correlation coefficient ranges from -1 to +1, where a value of +1 indicates a trend that consistently increases and never decreases with the opposite being true for a value of -1. A value of 0 indicates no consistent trend.

Observed changes in climate extremes

Daily weather data provided by the South African Weather Service for five stations in Namaqualand were used to analyse trends in temperature and rainfall extremes (Table 1). The weather station data were subjected to quality control to remove any erroneous values. The analysis was conducted using the RClimdex package (Version 1.1). This software package (available for download from http://etccdi.pacificclimate.org/software.shtml) has been used in a number of regional[1] and national studies[46,47] to detect changes in climate extremes. The software is capable of computing 27 core indices,[48] but only those temperature and rainfall indices relevant to Namaqualand were selected for this study (Tables 2 and 3). The base period used in the analysis is normally 1971–2000[1,46] but because the weather station data for Namaqualand only begans in the 1980s, the base period of 1980–2000 was used in the analysis. Linear trends were calculated using the least-squares method for each temperature and rainfall index. The trends were tested for significance at the 95% confidence level.

Table 2: Description of the temperature indices utilised in this study[1]

Index	Description	Definition	Units
TXx	Hottest day	Monthly highest TX	°C
TXn	Coolest day	Monthly lowest TX	°C
TX90P	Hot day frequency	Percentage of days when TX 90th percentile of the baseline average	%
TX10P	Cool day frequency	Percentage of days when TX>10th percentile of the baseline average	%
SU25	Hot days	Annual count when TX>25 °C	days
WSDI	Warm spell	Annual count days with at least 6 consecutive days when TX>90th percentile of the baseline average	days
TNn	Coolest day	Monthly lowest TN	°C
TNx	Hottest night	Monthly highest TN	°C
TN10P	Cool night frequency	Percentage of days when TN>10th percentile of the baseline average	%
TN90P	Hot night frequency	Percentage of days when TN>90th percentile of the baseline average	%
CSDI	Cold spell	Annual count days with at least 6 consecutive days when TN<10th percentile of the baseline average	days
TR20	Warm nights	Annual count when TN>20 °C	days
DTR	Diurnal temperature range	Monthly mean difference between TX and TN	°C

TX, daily maximum temperature; TN, daily minimum temperature

Table 3: Description of the rainfall indices utilised in this study[1]

Index	Description	Definition	Units
R95p	Very wet day	Annual total rainfall when RR>95th percentile of the baseline average	mm
CDD	Consecutive dry days	Maximum number of consecutive dry days	days
CWD	Consecutive wet days	Maximum number of consecutive wet days	days
R10mm	Heavy rainfall days	Annual count of days when RR>= 10 mm	days
R20mm	Very heavy rainfall days	Annual count of days when RR>= 20 mm	days
Rx5day	Maximum 5 day rainfall	Annual maximum consecutive 5 day rainfall	mm
SDII	Simple daily intensity index	Average rainfall on wet days	mm

RR, daily rainfall rate

Results

Observed changes in climate

Temperature

Linear trend analysis of the CRU TS 3.1 data set for Namaqualand revealed that annual minimum and maximum temperature have increased at a rate of 0.14 °C per decade ($p<0.0001$) and 0.11 °C per decade ($p<0.0001$) respectively between 1901 and 2009. It is clear from this data set that minimum temperatures are increasing at a faster rate than maximum temperatures and that the rate of warming has increased over the last two decades (Figure 2). Minimum temperatures for the period 1970–2009 were 0.36 °C warmer than the 1961–1990 mean whereas maximum temperatures for the same period were only 0.16 °C warmer than the 1961–1990 mean. The highest annual minimum temperature of the series was in 2009, with a temperature of 1.27 °C above the 1961–1990 mean. The highest annual maximum temperature of the series was for 1945, with a temperature of 0.96 °C above the 1961–1990 mean. The second highest annual maximum temperature was for 1999, with a temperature of 0.9 °C above the 1961–1990 mean.

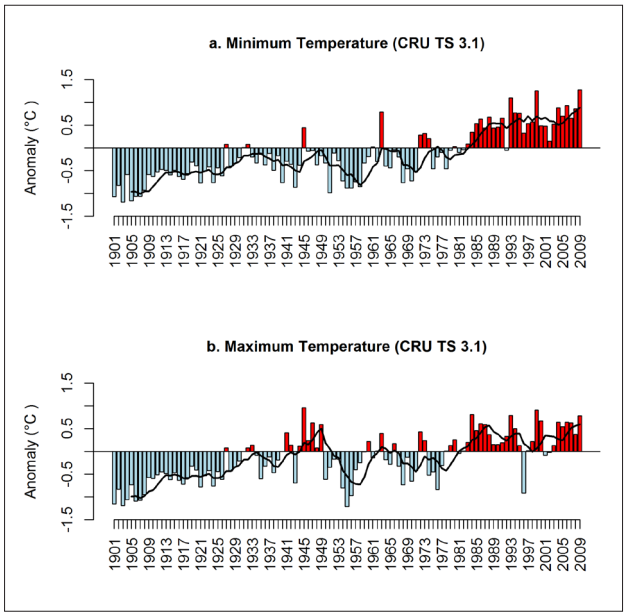


Figure 2: Annual (a) minimum and (b) maximum temperature anomalies (°C) for Namaqualand (1901–2009) based on the CRU TS 3.1 data set. Red represents positive anomaly and blue a negative anomaly in temperature with respect to the long-term average climatology (1961–1990 mean). The 5-year moving average (black line) is also shown.

A spatial pattern in temperature trends is evident (Figure 3 and Figure 4), with the greatest warming in minimum temperature being observed in the north-eastern parts of Namaqualand and the greatest warming in maximum temperature being observed in the south-eastern parts of Namaqualand towards the interior of the country. Trends in annual maximum temperature derived from the weather station data indicate a significant increase at the Port Nolloth and Vredendal stations of 0.3 °C and 0.1 °C per decade respectively (Supplementary Figure 1). Trends in minimum temperature were less consistent, with a significantly increasing trend observed for Pofadder, but a significantly decreasing trend observed for Springbok (Supplementary figure 2).

Temperature trends across seasons are inconsistent with slightly larger warming in spring (SON) and autumn (MAM) compared with the other seasons (Figures 3 and 4). Overall, areas further inland have experienced a higher rate of change than those areas along the coastline. As with the annual trends, the rate of increase in seasonal minimum temperatures is greater than that of maximum temperatures.

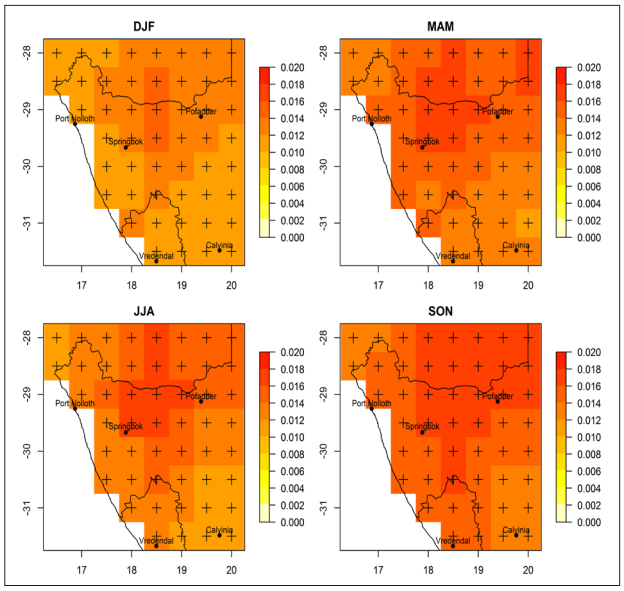


Figure 3: Observed seasonal minimum temperature trends (°C per year) according to the Theil-Sen trend analysis for summer (DJF), autumn (MAM), winter (JJA) and spring (SON) based on the CRU TS 3.1 data set (1901–2009). Statistically significant trends (95th confidence level) are stippled.

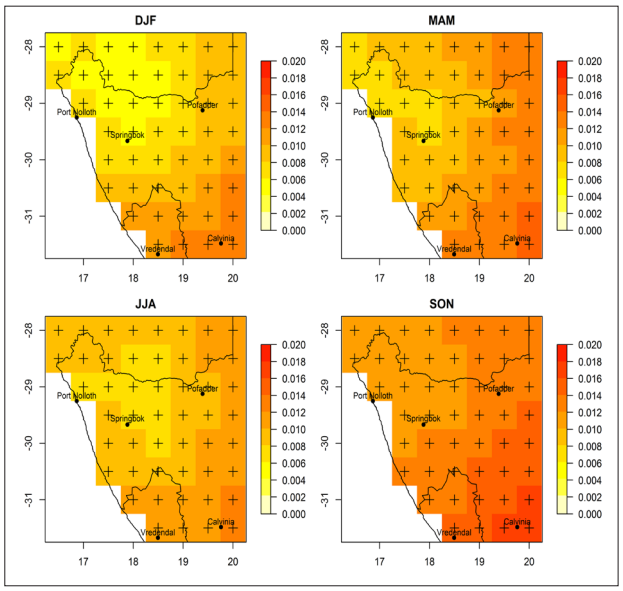


Figure 4: Observed seasonal maximum temperature trends (°C per year) according to the Theil-Sen trend analysis for summer (DJF), autumn (MAM), winter (JJA) and spring (SON) based on the CRU TS 3.1 data set (1901–2009). Statistically significant trends (95th confidence level) are stippled.

Rainfall

Based on the linear trend of mean annual rainfall derived CRU TS 3.1 data set from 1901 to 2009, there is no clear evidence for a significant change in rainfall over Namaqualand. Furthermore, the annual rainfall time series for each the weather stations demonstrates no significant change in rainfall over the last 30 years (Supplementary figure 3). The rainfall time series (Figure 5) is characterised by strong inter-annual variability with periods of above and below average rainfall, for example 1973–1976 and 2002–2004 respectively. There was no significant trend in the coefficient of variation (Supplementary figure 4), suggesting that the inter-annual rainfall variability has remained relatively constant throughout the period of analysis.

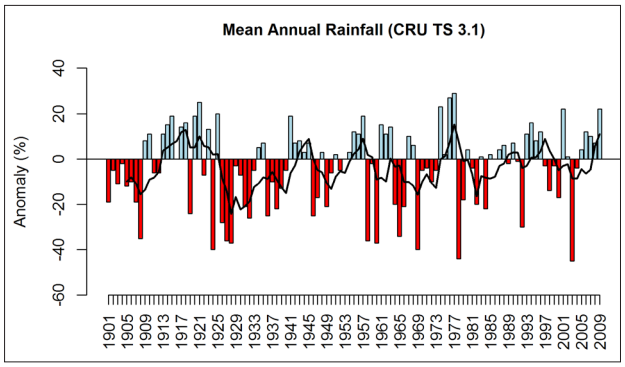


Figure 5: Annual rainfall anomalies (%) for Namaqualand (1901–2009) based on CRU TS 3.1 data set. Blue represents positive anomaly and red a negative anomaly with respect to the long-term average climatology (1961–1990 mean). The 5 year moving average (black line) is also given.

Seasonal trends in rainfall (Figure 6) do demonstrate some spatial patterns of change over Namaqualand. Rainfall during the main rainfall season (JJA) decreased over the eastern parts of Namaqualand and during spring (SON), rainfall decreased over the central and southern parts. These patterns may however be considered negligible because of the insignificance of the trends. The only statistically significant trends were observed in summer (DJF) over Port Nolloth, which has experienced a decrease, and over the Pofadder region which experienced an increase in rainfall. Additional analysis of the CRU TS 3.1 rainfall data set indicated that there was no significant trend in the rainfall totals over the main vegetation growing season (June–Oct)[49], (Supplementary Figure 5) and there was no evidence for shift in the seasonality or month of peak rainfall.

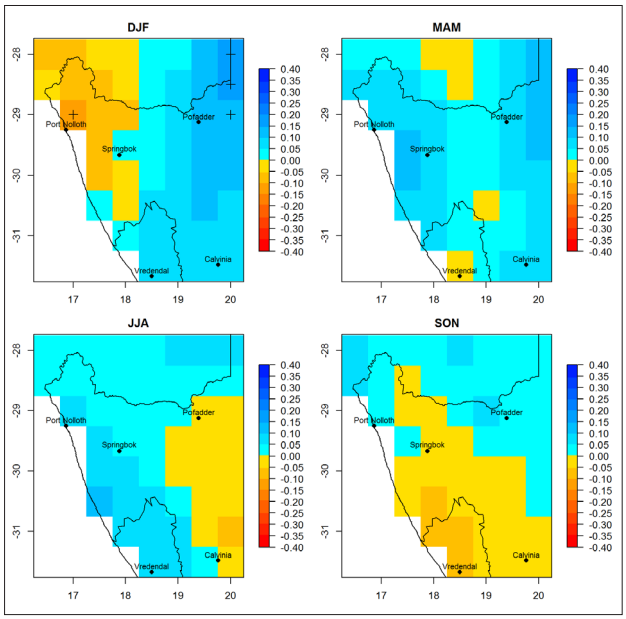


Figure 6: Observed trends in seasonal rainfall amounts according to the Mann-Kendall test for summer (DJF), autumn (MAM), winter (JJA) and spring (SON) based on the CRU TS 3.1 data set (1901–2009). The value represents the direction and the relative strength of the trend. Red colours indicate a negative trend and blue colours indicate positive trend. Statistically significant trends (95th confidence level) are stippled.

Evapotranspiration

Evidence from the analysis of MOD16 ET from 2000 to 2010 indicates that changes in evapotranspiration have not been uniform across Namaqualand. During spring (SON) and summer (DJF), the trend in evapotranspiration was characterised by a steepening inland-coastal gradient while areas along the coastline experienced an increase in evapotranspiration of up to 30 mm per decade over the period (Figure 7). This trend is most pronounced and statistically significant during spring along the coastline. During autumn (MAM), areas to the north of Namaqualand around Pofadder experienced a significant increase in evapotranspiration. During winter (JJA) there was a reduction in evapotranspiration over the entire region with significant trends observed over the escarpment and higher altitude regions of Springbok and Calvinia.

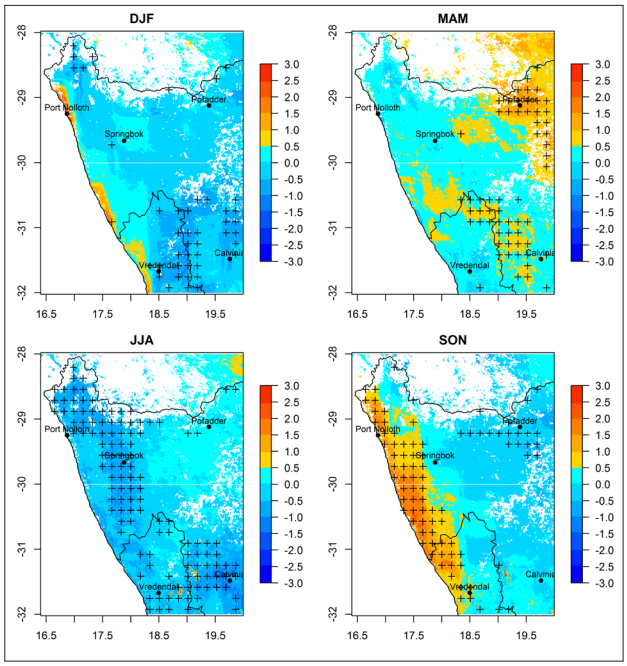


Figure 7: Observed trends in seasonal evapotranspiration (mm per season) according to the Theil-Sen trend analysis for summer (DJF), autumn (MAM), winter (JJA) and spring (SON) based on the CRU TS 3.1 data set (1901–2009). Blue colours indicate a decreasing trend, red colours indicate an increasing trend, and white indicates pixels with no data. Statistically significant trends (95th confidence level) are stippled.

Observed changes in climate extremes

The analysis of the weather station data indicate that the occurrence of warm extremes (Table 4) has increased and that the occurrence of cold extremes (Table 5) has decreased over Namaqualand. This is clearly demonstrated by the observed increase in the number of hot days (TX90P) and the corresponding decrease in the number of cool days (TX10P) for each weather station. Furthermore, all five of the weather stations demonstrated an increase in warm spell duration (WSDI), with those for Pofadder and Springbok being statistically significant. The duration of warm spells increased by 2 and 5 days per decade ($p<0.05$) at the Pofadder and Springbok stations respectively.

For the Vredendal weather station, the monthly highest maximum temperature, the frequency of hot days as well as the number of days above 25 °C all increased significantly. The number of hot days increased by 2.5 days ($p<0.05$) and the number of hot days above 25 °C by 12 days per decade ($p<0.05$) at this station.

For Pofadder, the frequency of hot nights (TN90P) increased by 1.1 days per decade ($p<0.05$) and the frequency of warm nights (TR20) by 0.7 days ($p<0.05$), corresponding with a decrease in the frequency of cool nights (TN10P) of 1.7 days ($p<0.05$). This resulted in a reduction in the diurnal temperature range (DTR) at this station. For the other weather stations, the occurrence of hot nights (TN90P) decreased, resulting in an increase in the DTR most notably for the Springbok and Vredendal weather stations.

Table 4: Trends in hot extreme indices derived from RClimDex expressed as the change per year. Trends that are significant at the 95% confidence level are indicted by an asterisk (*)

Weather station	Hottest day (TXx)	Hottest night (TNx)	Hot day frequency (TX90P)	Hot night frequency (TN90P)	Hot days (SU25)	Warm Spell (WSDI)	Warm nights (TR20)
Calvinia	-0.001	0.025	0.15	-0.006	0.416	0.15	0.039
Pofadder	0.004	0.009	0.115	0.106*	0.167	0.202*	0.716*
Port Nolloth	0.036	-0.032	0.176	-0.181	0.098	0.26	-0.024
Springbok	-0.005	-0.015	0.132	-0.221*	0.376	0.542*	-0.17
Vredendal	0.111*	0.028	0.25*	-0.044	1.222*	0.109	0.089

Table 5: Trends in cold extreme indices derived from RClimDex expressed as the change per year. Trends that are significant at the 95% confidence level are indicted by an asterisk (*)

Weather station	Coolest day (TNn)	Coolest day (TXn)	Cool night frequency (TN10P)	Cool day frequency (TX10P)	Cold Spell (CSDI)	Diurnal temperature range (DTR)
Calvinia	0.008	0.043	0.003	-0.101	0.08	0.024
Pofadder	-0.01	0.015	-0.167*	-0.114*	0.027	-0.006
Port Nolloth	0.022	-0.004	0.001	-0.132	0.244	0.054
Springbok	-0.002	0.03	0.054	-0.094	0.051	0.056*
Vredendal	0.007	0.018	0.045	-0.269*	0.045	0.069*

In terms of rainfall extremes (Table 6), the only significant trend observed was for Port Nolloth, that experienced a decline in the maximum consecutive 5 day rainfall (Rx5day), ($p<0.05$). The trends for the other weather stations were weaker and differed considerably in both magnitude and direction thus providing little evidence for a change in rainfall extremes.

Discussion

Using a high-resolution (0.5°x0.5°) gridded data set (CRU TS 3.1), weather station data and MOD16 evapotranspiration data, an updated analysis of changes in climate for the Namaqualand region of South Africa was provided. There is strong evidence from the CRU TS 3.1 high-resolution gridded data set to suggest that temperatures in Namaqualand have been increasing over the last century, and that the rate of warming has been increasing – most notably in the last two decades. The trends presented are consistent with detected increases in global regional temperatures.[2,3,5,9,50] Projections of future temperature change for South Africa[9] indicate that temperatures are expected to continue to increase.

According to the analysis of CRU TS 3.1 data set, minimum temperatures are increasing at a faster rate than maximum temperatures. As demonstrated in other studies,[50-52] this has resulted in a decrease in the DTR for many parts of the globe. The results from the analysis of the weather station data on the other hand, demonstrated inconsistent trends in minimum temperature with DTR decreasing at Pofadder but increasing at the other 4 stations. The difference between the two data sets could be attributed to weak correlations, especially for the measurement of minimum temperature, at the Springbok and Port Nolloth stations. However, two recent studies in South Africa[8,46] found no clear consistent pattern with regards to changes in DTR and suggested that this topic requires further exploration and research.

Data from the five weather stations studied suggest that the occurrence of hot extremes has increased over Namaqualand, most notably for the Pofadder weather station, which is located more towards the interior of the country. There has been a clear increase in the frequency of hot days and a corresponding decrease in the frequency of cool days at each weather station which could indicate a shift in the statistical distribution of maximum temperatures in the region. The lack of statistically significant trends for the weather stations could be a result of the relatively short time series analysed (1982–2014). Noting this limitation, the direction of the trends presented here are in line with regional studies.[1,8,9,46]

Table 6: Trends in rainfall indices derived from RClimDex expressed as the change per year. Trends that are significant at the 95% confidence level are indicted by an asterisk (*)

Weather station	Very wet day (R95p)	Consecutive dry days (CDD)	Consecutive wet days (CWD)	Heavy rainfall days (R10mm)	Very heavy rainfall days (R20mm)	Maximum 5 day rainfall (Rx5day)	Simple daily intensity index (SDII)
Calvinia	-0.524	-1.084*	0.011	0.029	-0.005	-0.02	-0.03
Pofadder	0.518	0.434	0.01	0.02	0.025	0.193	0.008
Port Nolloth	-0.316	0.772	-0.022	0.009	-0.01	-0.388*	-0.044
Springbok	1.073	-0.265	-0.043	0.025	0.049	0.375	0.011
Vredendal	0.245	1.136	-0.005	-0.014	0.027	0.36	0.012

Changes in rainfall are typically harder to detect because rainfall varies depending on location and from year to year.[6,53] From the results presented here, no clear evidence exists for a significant change in mean annual rainfall, and the rainfall time series remains dominated by oscillating wet and dry conditions. This is consistent with previous studies[19,54] which found no evidence for a significant change in mean annual rainfall over the Succulent Karoo. More recent studies for South Africa have detected decreases in rainfall and the number of rainfall days over parts of the country[8] but overall, the trends are weak and non-significant.[9]

We did however demonstrate that rainfall has increased in summer (DJF) over the north-eastern region of Namaqualand. This could be attributed to either more favourable synoptic scale dynamics or local-scale land surface processes during these months.[18] The reduction in winter rainfall (JJA), the main rainfall season in Namaqualand, was also noted by MacKellar et al.[18] but the trend presented here is insignificant, thus the change could be considered negligible. An increase in summer rainfall could promote the development of more favourable conditions for C4 grasses which could threaten the persistence of endemic species.[14]

The strong inter-annual variability in rainfall in Namaqualand has been noted in previous studies[19,54] but no trend in the frequency of drought and wet periods has been observed.[17] Recent evidence suggests that El Nino Southern Oscillation (ENSO) modulates rainfall in the region with El Nino (La Nina) years being associated with higher (lower) than normal rainfall amounts in May, June and July.[55] During ENSO events, the rain-bearing systems are larger in extent and are located further north, while during La Nina Southern Oscillation (LNSO), events are smaller and located further south.[55]

The latest projections from downscaled climate models indicate that rainfall is expected to decrease over Namaqualand in the future.[9,56] Although these downscaled models are becoming increasingly sophisticated, there is still some uncertainty in rainfall projections at local scale.[2,57] This mismatch between observed and predicted changes for Namaqualand has important implications for how changes in rainfall are perceived by on-the-ground managers as well as the kinds of management actions and adaptation response measures that are prioritised.

In Namaqualand, dewfalls and coastal fog are significant sources of moisture for the region and have been shown at times to exceed the amount rainfall during the wet season.[58] Assessments of changes in these two key variables are required in order to have an improved understanding of future changes in water availability for the region. Very little is known about the variability and long term changes in coastal fog off the west coast of southern Africa but efforts are currently underway through a long term research programme, FogLife[59], to address this gap.

As data on dewfalls and coastal fog are limited, we extended our analysis to include an assessment of trends in evapotranspiration. There were some clear spatial patterns in the trends in evapotranspiration over Namaqualand which appear to be influenced primarily by topography. Areas along the coastline and areas in the north-east of Namaqualand experienced the greatest increase in evapotranspiration, most notably during spring (SON) and autumn (MAM) respectively. The strongest warming was noted during spring and autumn and consequently, this observed increase in evapotranspiration is likely to be a direct result of an increase in temperature. The marked increase over the coastline in spring could however be indicative of a change in fog along the coast, but this finding requires more detailed analysis and monitoring of data on coastal fog. The higher altitude regions, on the other hand, have experienced a reduction in evapotranspiration, most notably in winter (JJA). A decrease over these areas could be explained by decreased wind speed and/or decreased solar radiation receipt as a result of increased cloud cover and atmospheric aerosol content.[60] There are no previous studies on the trends in evapotranspiration for the Succulent Karoo or the Namaqualand region. In one of the only related studies, Hoffman et al.[61] found that pan evaporation declined significantly at an average rate of 9.1 mm per annum at 16 weather stations located in the Cape Floristic Region of South Africa to the south of Namaqualand as a result of a decline in wind run.

The exceptional diversity of the Namaqualand has been attributed to firstly, the low but relatively predictable annual rainfall amounts and secondly, the moderate temperature regime throughout the year.[32,36] The increase in temperature combined with the observed increases in evapotranspiration is of significant concern for Namaqualand especially as scenarios of future climate change project even stronger changes in the next decade. The succulent flora of Namaqualand evolved during a period when it was cooler and probably also wetter[14,62] and thus higher temperatures are not likely to have been previously experienced in the region and could create conditions unsuitable for many species. Even if annual rainfall totals and coastal fog remain unchanged, future temperature increase will exert water stress on the plants in the region.[58,62] Furthermore, the continued increase in temperature for the region has considerable implications for livestock heat stress as well as for other agricultural sectors, including the viability of certain pests and pathogens.[63]

Studies, mostly in the northern hemisphere, have shown that temperature changes have had a large impact on the timing of phenological events.[64] The onset of flowering in Namaqualand generally occurs in spring and is triggered by a change in temperature,[65] while the amount of rainfall during winter is usually responsible for numbers of flowers and duration of flowering.[66] Persistent high temperatures, particularly towards the end of the growing season have shown to have a negative effect on plants in Namaqualand.[66] The combined effects of increased water stress and high temperatures may pose significant challenges for Namaqualand.

Continued long-term monitoring of both changes in climate and vegetation in Namaqualand is essential in order to fully understand the extent of the impacts of increased temperature and changing patterns of water availability. The results presented in this paper provide a foundation for more detailed climate change assessments for the region as well as motivation for the development of adaptation responses and the continued investment in existing regional programmes. The gaps in weather information identified here, such as the lack coastal fog and dewfall data, need to be urgently addressed through a better network of weather stations as well as improved coordination across monitoring systems. Other key priorities for future research include detailed studies of the changes in diurnal temperature ranges on a broader regional scale and the continued monitoring of both climate observations and species-level changes in Namaqualand in order to test and refine climate change projections.

Authors' contributions

This paper is based on the work conducted by C.L.D as part of her MSc thesis at the University of Cape Town. M.T.H was the primary supervisor and W.R, the secondary supervisor.

References

1. New M, Hewitson B, Stephenson DB, Tsiga A, Kruger A, Manhique A, et al. Evidence of trends in daily climate extremes over southern and west Africa. J Geophys Res. 2006;111(7):D14102. http://dx.doi.org/10.1029/2005JD006289
2. Stocker T, Qin D, Platner G. Climate Change 2013: The Physical Science Basis. Working Group I. Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers. Cambridge: Cambridge University Press; 2013.
3. Mason S, Jury M. Climatic variability and change over southern Africa: A reflection on underlying processes. Prog Phys Geogr. 1997;21(1):23–50. http://dx.doi.org/10.1177/030913339702100103
4. Mason SJ, Waylen PR, Mimmack GM, Rajaratnam B, Harrison JM. Changes in extreme rainfall events in South Africa. Clim Change. 1999;41(2):249–257. http://dx.doi.org/10.1023/A:1005450924499
5. Hulme M, Doherty R, Ngara T, New M, Lister D. African climate change: 1900–2100. Clim Res. 2001;17(2):145–168. http://dx.doi.org/10.3354/cr017145
6. Fauchereau N, Trzaska S, Rouault M, Richard Y. Rainfall variability and changes in southern Africa during the 20th century in the global warming context. Nat Hazards. 2003;29(2):139–154. http://dx.doi.org/10.1023/A:1023630924100

7. Kruger A, Shongwe S. Temperature trends in South Africa: 1960–2003. Int J Climatol. 2004;24(15):1929–1945. http://dx.doi.org/10.1002/joc.1096

8. MacKellar N, New M, Jack C. Observed and modelled trends in rainfall and temperature for South Africa: 1960–2010. S Afr J Sci. 2014;110(7/8): Art. #2013-0353, 13 pages. http://dx.doi.org/10.1590/sajs.2014/20130353

9. Department of Environmental Affairs. Long-Term Adaptation Scenarios Flagship Research Programme (LTAS) for South Africa: Climate trends and scenarios for South Africa. Pretoria: Department of Environmental Affairs; 2013. Available from: http://www.sanbi.org/sites/default/files/documents/documents/ltassummary-policy-makers2013high-res.pdf

10. Scholes RJ, Biggs RA. Ecosystem services in southern Africa: A region assessment. The regional-scale component of the southern African millennium ecosystem assessment. Pretoria: Council for Scientific and Industrial Research; 2004.

11. Gitay H, Suárez A, Watson RT, Dokken DJ. Climate change and biodiversity: Intergovernmental panel on climate change (IPCC) technical paper V. Geneva: IPCC; 2002.

12. Benjaminsen TA, Rohde R, Sjaastad E, Wisborg P, Lebert T. Land reform, range ecology, and carrying capacities in Namaqualand, South Africa. Ann Assoc Am Geogr. 2006;96(3):524–540. http://dx.doi.org/10.1111/j.1467-8306.2006.00704.x

13. Cousins B, Hoffman M, Allsopp N, Rohde R. A synthesis of sociological and biological perspectives on sustainable land use in Namaqualand. J Arid Environ. 2007;70(4):834–846. http://dx.doi.org/10.1016/j.jaridenv.2007.04.002

14. Midgley G, Thuiller W. Potential vulnerability of Namaqualand plant diversity to anthropogenic climate change. J Arid Environ. 2007;70(4):615–628. http://dx.doi.org/10.1016/j.jaridenv.2006.11.020

15. Rutherford MC, Midgley GF, Bond WJ, Powrie LW, Roberts R, Allsopp J. Plant biodiversity: Vulnerability and adaptation assessment. In: Kiker G, editor. Climate change impacts in southern Africa. Report to the National Climate Change Committee. Pretoria: Department of Environment Affairs and Tourism; 2000.

16. Broennimann O, Thuiller W, Hughes G, Midgley GF, Alkemade JR, Guisan A. Do geographic distribution, niche property and life form explain plants' vulnerability to global change? Global Change Biol. 2006;12(6):1079–1093. http://dx.doi.org/10.1111/j.1365-2486.2006.01157.x

17. Kelso C, Vogel C. The climate of Namaqualand in the nineteenth century. Clim Change. 2007;83(3):357–380. http://dx.doi.org/10.1007/s10584-007-9264-1

18. MacKellar N, Hewitson B, Tadross M. Namaqualand's climate: Recent historical changes and future scenarios. J Arid Environ. 2007;70(4):604–614. http://dx.doi.org/10.1016/j.jaridenv.2006.03.024

19. Hoffman M, Carrick P, Gillson L, West A. Drought, climate change and vegetation response in the Succulent Karoo, South Africa. S Afr J Sci. 2009;105(1/2):54–60. http://dx.doi.org/10.1590/S0038-23532009000100021

20. Chase BM, Meadows ME. Late Quaternary dynamics of southern Africa's winter rainfall zone. Earth-Sci Rev. 2007;84(3):103–138. http://dx.doi.org/10.1016/j.earscirev.2007.06.002

21. Benito G, Thorndycraft V, Rico M, Sánchez-Moya Y, Sopeña A, Botero B, et al. Hydrological response of a dryland ephemeral river to southern African climatic variability during the last millennium. Quatern Res. 2011;75(3):471–482. http://dx.doi.org/10.1016/j.yqres.2011.01.004

22. Benito G, Botero BA, Thorndycraft VR, Rico M, Sánchez-Moya Y, Sopeña A, et al. Rainfall-runoff modelling and palaeoflood hydrology applied to reconstruct centennial scale records of flooding and aquifer recharge in ungauged ephemeral rivers. Hydrol Earth Syst Sci. 2010;7(6):9631–9660.

23. Weldeab S, Stuut J, Schneider R, Siebel W. Holocene climate variability in the winter rainfall zone of South Africa. Clim Past. 2013;9:2309–2356. http://dx.doi.org/10.5194/cpd-9-2309-2013

24. Bourne A, Donatti C, Holness S, Midgley G. Climate change vulnerability assessment for the Namakwa District Municipality. Cape Town: Conservation South Africa; 2012.

25. Namakwa District Municipality. Integrated development plan 2012–2016. Springbok: Namakwa District Municipality; 2012.

26. New M, Lister D, Hulme M, Makin I. A high-resolution data set of surface climate over global land areas. Clim Res. 2002;21(1):1–25. http://dx.doi.org/10.3354/cr021001

27. Hewitson BC, Crane RG. Gridded area-averaged daily precipitation via conditional interpolation. J Clim. 2005;18(1):41–57. http://dx.doi.org/10.1175/JCLI3246.1

28. Mitchell TD, Jones PD. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int J Climatol. 2005;25(6):693–712. http://dx.doi.org/10.1002/joc.1181

29. Harris I, Jones P, Osborn T, Lister D. Updated high-resolution grids of monthly climatic observations – the CRU TS3. 10 Dataset. Int J Climatol. 2013;34:623–642. http://dx.doi.org/10.1002/joc.3711

30. Cowling R, Hilton-Taylor C. Plant biogeography, endemism and diversity. In: Dean WRJ, Milton SJ, editors. The Karoo: Ecological patterns and processes. Cambridge: University Press; 1999. p. 42–56. http://dx.doi.org/10.1017/CBO9780511541988.007

31. Desmet P. Namaqualand: A brief overview of the physical and floristic environment. J Arid Environ. 2007;70(4):570–587. http://dx.doi.org/10.1016/j.jaridenv.2006.11.019

32. Cowling R, Esler K, Rundel P. Namaqualand, South Africa – an overview of a unique winter-rainfall desert ecosystem. Plant Ecol. 1999;142(1):3–21. http://dx.doi.org/10.1023/A:1009831308074

33. Rohde RF, Benjaminsen TA, Hoffman M. Land reform in Namaqualand: – Poverty alleviation, stepping stones and 'economic units'. Land reform and agrarian change in southern Africa series; no. 16. Bellville: Programme for Land and Agrarian Studies; 2001.

34. Hoffman MT, Cousins C, Meyer T, Petersen A, Hendriks H. Historical and contemporary agricultural land use and the desertification of the Karoo. In: Dean WRJ, Milton S, editors. The Karoo: Ecological patterns and processes. Cambridge: Cambridge University Press; 1999. http://dx.doi.org/10.1017/CBO9780511541988.022

35. Tyson PD, Preston-Whyte RA. The weather and climate of southern Africa. Cape Town: Oxford University Press; 2000.

36. Desmet P, Cowling R. The climate of the Karoo - a functional approach. In: Dean WRJ, Milton S, editors. The Karoo: Ecological patterns and processes. Cambridge: University press; 1999. p. 3–16. http://dx.doi.org/10.1017/CBO9780511541988.004

37. Driver A, Maze K. The Succulent Karoo ecosystem plan (SKEP): An introduction to SKEP. Veld Flora 2002; 88(1):12–13.

38. Hulme M. A 1951–1980 global land precipitation climatology for the evaluation of general circulation models. Clim Dyn. 1992;7(2):57–72. http://dx.doi.org/10.1007/BF00209609

39. Houghton JT, Ding Y, Griggs DJ, Noguer M, Van der Linden PJ, Dai X, et al. Clim Change. 2001: The Scientific Basis. Cambridge: Cambridge University Press; 2001.

40. Mu Q, Zhao M, Running SW. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens Environ. 2011;115(8):1781–1800. http://dx.doi.org/10.1016/j.rse.2011.02.019

41. Monteith J. Evaporation and environment. Sym Soc Exp Biol. 1965;19:205–223.

42. A rank-invariant method of linear and polynomial regression analysis, Part 3. P K Ned Akad A Math; 1950.

43. Sen PK. Estimates of the regression coefficient based on Kendall's tau. J Am Stat Assoc. 1968;63(324):1379–1389. http://dx.doi.org/10.1080/01621459.1968.10480934

44. Neeti N, Eastman JR. A contextual Mann-Kendall approach for the assessment of trend significance in image time series. Transac GIS. 2011;15(5):599–611. http://dx.doi.org/10.1111/j.1467-9671.2011.01280.x

45. Kendall MG. A new measure of rank correlation. Biometrika. 1938;30(1/2):81–93. http://dx.doi.org/10.1093/biomet/30.1-2.81

46. Kruger A, Sekele S. Trends in extreme temperature indices in South Africa: 1962–2009. Int J Climatol. 2013;33(3):661–676. http://dx.doi.org/10.1002/joc.3455

47. Kruger A. Observed trends in daily precipitation indices in South Africa: 1910–2004. Int J Climatol. 2006;26(15):2275–2285. http://dx.doi.org/10.1002/joc.1368

48. Zhang X, Yang F. RClimDex (1.0) User Guide. Downsview: Environment Canada, Climate Research Branch; 2004.

49. Fox S, Hoffman M, Hoare D. The phenological pattern of vegetation in Namaqualand, South Africa and its climatic correlates using NOAA-AVHRR NDVI data. S Afr Geogr J. 2005;87(2):85–94. http://dx.doi.org/10.1080/03736245.2005.9713831

50. Easterling DR, Horton B, Jones PD, Peterson TC, Karl TR, Parker DE, et al. Maximum and minimum temperature trends for the globe. Science. 1997;277(5324):364–367. http://dx.doi.org/10.1126/science.277.5324.364

51. Karl TR, Jones PD, Knight RW, Kukla G, Plummer N, Razuvayev V, et al. A new perspective on recent global warming: Asymmetric trends of daily maximum and minimum temperature. Bull Amer Meteor Soc. 1993;74:1007–1023. http://dx.doi.org/10.1175/1520-0477(1993)074<1007:ANPORG>2.0.CO;2

52. Blunden J, Arndt DS. State of the Climate in 2011. Bull Amer Meteor Soc. 2012;93:S1–S282. http://dx.doi.org/10.1175/2012BAMSStateoftheClimate.1

53. Hoffman T, Vogel C. Climate change impacts on African rangelands. Rangelands. 2008;30(3):12–17. http://dx.doi.org/10.2111/1551-501X(2008)30[12:CCIOAR]2.0.CO;2

54. Haensler A, Hagemann S, Jacob D. Climate history of Namibia and western South Africa. In: Jürgens N, Schmiedel U, Hoffman MT, editors. Biodiversity in southern Africa: Patterns and Processes at Regional Scale Vol 2. Göttingen & Windhoek: Klaus Hess Publishers; 2010. p. 2.

55. Philippon N, Rouault M, Richard Y, Favre A. The influence of ENSO on winter rainfall in South Africa. Int J Climatol. 2011;32(15):2333–2347. http://dx.doi.org/10.1002/joc.3403

56. Engelbrecht F, McGregor J, Engelbrecht C. Dynamics of the Conformal-Cubic Atmospheric Model projected climate-change signal over southern Africa. Int J Climatol. 2009;29(7):1013–1033. http://dx.doi.org/10.1002/joc.1742

57. Knutti R, Sedlácek J. Robustness and uncertainties in the new CMIP5 climate model projections. Nature Clim Change. 2013;3(4):369–373. http://dx.doi.org/10.1038/nclimate1716

58. Schmiedel U, Dengler J, Etzold S. Vegetation dynamics of endemic-rich quartz fields in the Succulent Karoo, South Africa, in response to recent climatic trends. J Veg Sci. 2012;23(2):292–303. http://dx.doi.org/10.1111/j.1654-1103.2011.01346.x

59. FogLife Colloquium. FogLife Colloquium Proceedings, 2014 Nov 11–13; Gobabeb, Namibia. Available from: http://www.gobabebtrc.org/images/2014_FogLife_Colloquium_Proceedings.pdf;

60. Eamus D, Palmer AR. Is climate change a possible explanation for woody thickening in arid and semi-arid regions? Int J Ecol. 2008;2007:1–5. http://dx.doi.org/10.1155/2007/37364

61. Hoffman MT, Cramer MD, Gillson L, Wallace M. Pan evaporation and wind run decline in the Cape Floristic Region of South Africa (1974–2005): Implications for vegetation responses to climate change. Clim Change. 2011;109(3):437–452. http://dx.doi.org/10.1007/s10584-011-0030-z

62. Musil C, Van Heerden P, Cilliers C, Schmiedel U. Mild experimental climate warming induces metabolic impairment and massive mortalities in southern African quartz field succulents. Environ Exp Bot. 2009;66(1):79–87. http://dx.doi.org/10.1016/j.envexpbot.2008.11.008

63. Archer E, Oettlé N, Louw R, Tadross M. 'Farming on the edge' in arid western South Africa: Climate change and agriculture in marginal environments. Geography. 2008;93:98–107.

64. Walther G, Post E, Convey P, Menzel A, Parmesan C, Beebee TJ, et al. Ecological responses to recent climate change. Nature. 2002;416(6879):389–395. http://dx.doi.org/10.1038/416389a

65. Van Rooyen M, Theron G, Grobbelaar N. Phenology of the vegetation in the Hester Malan Nature Reserve in the Namaqualand Broken Veld: 2. The therophyte population. J S Afr Botany. 1979;45:433–452.

66. Steyn H, Van Rooyen N, Van Rooyen M, Theron G. The phenology of Namaqualand ephemeral species. The effect of water stress. J Arid Environ. 1996;33(1):49–62. http://dx.doi.org/10.1006/jare.1996.0045

Note: This article includes supplementary material

Antimutagenic and antioxidant effects of a South African traditional formulation used as an immune booster

AUTHORS:
Mlungisi Ngcobo[1]
Nceba Gqaleni[2]
Victor Ndlovu[3]

AFFILIATIONS:
[1]Traditional Medicine Laboratory, School of Nursing and Public Health, College of Health Sciences, University of KwaZulu Natal, Durban, South Africa
[2]Department of Public Management and Economics, Durban University of Technology, Durban, South Africa
[3]Makhonya Natural Health Products (Pty) Ltd, Durban, South Africa

CORRESPONDENCE TO:
Nceba Gqaleni

EMAIL:
nceba5850@gmail.com

POSTAL ADDRESS:
Department of Public Management and Economics, Durban University of Technology, 79 Steve Biko Road, Durban 4001, South Africa

DATES:
Received: 21 Apr. 2015
Revised: 11 Aug. 2015
Accepted: 05 Oct. 2015

KEYWORDS:
traditional medicines; regulations; genotoxicity; antioxidants; immune cells

HOW TO CITE:
Ngcobo M, Gqaleni N, Ndlovu V. Antimutagenic and antioxidant effects of a South African traditional formulation used as an immune booster. S Afr J Sci. 2016;112(3/4), Art. #2015-0152, 7 pages. http://dx.doi.org/10.17159/sajs.2016/20150152

The traditional medicines sector in South Africa is still largely unregulated despite legislation aimed at regulating the practice being in place. The HIV and AIDS epidemic has fuelled demand for traditional medicines, with many patients consulting traditional health practitioners who offer different treatments, including herbal immune boosters. This study investigated the mutagenic and antioxidant effects of the widely sold herbal immune booster, *uMakhonya*®. The Ames test was used for analysis of the genototoxic effects while the adenosine triphosphate (ATP) assay was used to evaluate cell cytotoxicity in peripheral blood mononuclear cells (PBMCs) and THP-1 monocytes. To evaluate the antioxidant effects the malondialdehyde (MDA) quantification, the nitric oxide and 1,1-diphenyl-2-picryl-hydrazyl (DPPH) assays were used. *UMakhonya*® doses of up to 5000 μg/mL were not genotoxic in the Ames test. *UMakhonya*® was shown to induce dose-dependent cytotoxicity in both PBMCs and THP-1 cells with doses ranging from 500 μg/mL to 1000 μg/mL, showing significant ($p<0.05$) toxicity. *UMakhonya*® was able to significantly ($p<0.05$) reduce nitrite radicals at 100 μg/mL while lower doses were not effective when compared to samples stimulated by lipopolysaccharide only. Non-cytotoxic doses of *uMakhonya*® showed significant ($p<0.05$) lipid peroxide scavenging ability in supernatants while this scavenging ability was considerably reduced intracellularly. In the DPPH assay, when both *uMakhonya*® and ascorbic acid were reconstituted in buffered saline, the traditional herbal remedy showed better radical scavenging abilities. Therefore further studies on the genotoxicity of *uMakhonya*®, when metabolically activated, and its antioxidant effects in in-vivo models are warranted.

Introduction

Traditional healers have for centuries advocated the value of using a combination of herbal remedies, single extracts and combined extracts to switch on the body's defence mechanisms, self-healing and protective processes.[1] Herbal plants are mostly used for this purpose and are often referred to as bitter tonics, adaptogens, stimulants, immune stimulants, immune boosters and strengthening mixtures. These African tonic plants are used for specific outcomes such as reducing fatigue, improving general health (during or after illness), reducing stress and cleansing the blood. Some aphrodisiacs for men are also made from tonic plants.[2] Stress related bio-chemicals are known to play a significant role in immune suppression and mental interventions using traditional methods of healing and can play an important role in relieving the effects of such suppression.[3] Therefore the study of the relationship between the effects of immune stimulants on the inflammatory response and the possible induction of oxidative stress and mutagenicity is important.

As complementary and alternative medicine has become more popular and attractive in the developed world, there has been an increasing interest by the scientific community to study the safety, efficacy and the mechanism of action of multiple herbal medicine.[4] Most of the studied African herbal medicinal extracts are made from a single plant, while most of the traditional medicines are made up of a combination of two or more plants. Scientific information regarding plants used in African traditional medicine in the form of mixtures and their effect on human health or on genetic material is poorly understood. Of those traditional medicines that have been researched and studied, the medicinal plants have been shown to be safe and effective in improving the health status of patients and hence have warranted further research.[5] Therefore, the study of the relationship between the effects of immune stimulants on the inflammatory response and the possible induction of oxidative stress and mutagenicity is important. Cellular mechanisms and external factors involved in the production of oxidative stress include the inflammatory response, peroxidation of cell membrane lipids and pro-oxidant activities of toxins.[6] Consumption of herbal products has been linked to reduced risk of conditions such as cancer and cardiovascular disease.[7] These potential benefits were demonstrated in a recent study of traditional herbal preparations sold in South Africa which showed that these preparations possess high antioxidant potential, reverse transcriptase inhibition or acetylcholinesterase enzyme inhibitory activity greater than 50%.[7] These observed activities not only demonstrate the potential benefits of these herbal preparations but necessitate further research studies.

Plants are known to be a rich source of secondary metabolites. Of all current pharmaceutical products, 25–50% are derived from plants.[8] Compounds from plants could act as protective agents with respect to human carcinogenesis, acting against the initiation, promotion or progression stages of this process or, perhaps by destroying or blocking the DNA-damaging mutagens outside the cells, thus avoiding cell mutations.[9] Many naturally occurring compounds with antioxidant activity are known to protect cellular components from oxidative damage and prevent diseases.[10] A number of such compounds can activate the phase II detoxification enzymes, which can remove the toxic elements from the system.[11] This study investigated the mutagenic and antioxidant effects of a South African commercial traditional medicinal product, *uMakhonya*®, which is used as an immune tonic. This product is formulated by combining five different medicinal plants, including *Artemisia afra*, Menthol, *Psidium guajava* liquid extract, *Chondrus crispus*, and

Uncaria tomentosa. *Artemisia afra* is used extensively to treat respiratory ailments and fever, suggesting an ability to increase resistance.[2] Menthol is used in foods, topical therapeutic preparations, oral hygiene and dentifrice formulations, and tobacco products by virtue of its pleasant minty flavour and the cooling sensation it imparts when in contact with the skin or oral membranes.[12] *Psidium guajava* leaf, root, and bark extracts are used traditionally for the treatment of diarrhoea, leukorrhea, cholera, external ulcers and skin diseases.[13] Seaweed like *Chondrus crispus* contains bioactive substances like polysaccharides, proteins, lipids and polyphenols, with antibacterial, antiviral and antifungal properties.[14,15] *Uncaria tomentosa* (cat's claw) is a medicinal plant from the Amazon forest in South America used for treatment of a wide range of diseases, including arthritis, gastritis, osteoarthritis, diabetes, and cancer.[16] According to statistics provided by the owner of *uMakhonya*®, over 16 500 1 L containers of this immune tonic were sold in 2014.

Material and Methods

Ethical clearances

This study received ethical approval from the Biomedical Research Ethics Committee of the University of KwaZulu-Natal (Reference number: BE168/11). Human whole blood samples were kindly donated by the South African National Blood Service (SANBS) (Human Research Ethics Committee Certificate Number: 2012/07).

Materials

The THP-1 monocyte cells were a gift from Mr Saiyur Ramsugit of the discipline of Medical Microbiology, University of KwaZulu Natal. Roswell Park Memorial Institute (RPMI)-1640 medium with L-glutamine, foetal calf serum (FCS), penicillin-streptomycin-fungizone (PSF), L-glutamine, Opti-MEM, Histopaque 1077 and Hepes buffer were purchased from Lonza (Johannesburg, South Africa). The lipopolysaccharide (LPS) from *Salmonella typhosa*, Peptidogylcan from *Staphylococcus aureus*, Cyclosporine, 1,1-Diphenyl-2-picryl-hydrazyl (DPPH), Ascorbic acid, 4-nitroquinoline-1-oxide (4-NQO), phorbol 12-myristate 13-acetate (PMA) and Polymyxin B sulphate were purchased from Sigma Aldrich (St Louis, MO, USA). The Promega CellTiter-Glo™ Luminescent Cell Viability assay kit and the Griess Reagent System were purchased from Promega (Fitchburg, WI, USA). The OxiSelect™ TBARS Assay Kit (MDA Quantitation) kit was purchased from Cell Biolabs Inc (San Diego, CA, USA). The automated cell counter was from BioRad (Hercules, CA, USA). Glo-Max Modulus™ Microplate Luminometer was from Turner BioSystems (Sunnyvale, CA, USA). The colometric plate reader was from Zenyth200 (Cambridge, UK). All other reagents and equipment were purchased from standard commercial sources and were of the highest available purity analytical grade.

Methods

Preparation of the *uMakhonya*® formula

UMakhonya® formulation samples (Batch number: R0101, National Pharmaceutical Product Index (NAPPI) Code: 710345-001) were donated by the owner of the traditional medicine formulation. He was actively involved in the research process and gave the researchers a tour of the industrial plant where the product is formulated and manufactured. Samples of these plants were given to our research group as reference. The herbal plants are listed on the packaging of the formula when sold in supermarkets and each 5 mL of *uMakhonya*® contains *Artemisia afra* (90 mg), Menthol (8.5 mg), *Psidium guajava* liquid extract (0.3 mL), *Chondrus crispus* (0.08 mL), and *Uncaria tomentosa* (Cats-claw tincture, 0.06 mL) and the rest is water. The formulation is listed with the Medicines Control Council (MCC) of South Africa and exclusive rights are reserved. It is manufactured by UMakhonya Natural Health Products (Pty) Ltd (Pinetown, South Africa). At the manufacturing phase, the plants are mixed proportionally in a 25 L tank and then extracted using tap water by boiling the contents of the tank at 100 °C overnight. The extract is then cooled, filtered once with a steel sift followed by removal of finer particles using a sifting net and then packaged into 1 L or 5 L containers.

To prepare the extract for in-vitro studies, the liquid extract was further sterile filtered and then freeze-dried to powder. The powdered plant material was then reconstituted at 10 mg/mL in phosphate buffered saline (PBS) (pH 7.2) and this was further sterile filtered with 0.22 μm filters. Working concentrations ranging between 1000 μg/mL and 10 μg/mL were then made using complete culture media. Endotoxin contamination was measured using the Limulus Amebocyte Lysate QVL-1000™ (Lonza, USA) with a sensitivity of 0.1 endotoxin units per mL. Polymyxin B sulphate (10 μg/mL) was added to reduce the immunostimulatory effects resulting from endotoxin contamination.

Ames test

The bacterial strains used for the mutagenicity testing were the histidine-requiring *Salmonella typhimurium* tester strains TA98 (detects frameshift mutagens) and TA100 (detects mutagens that cause base-pair substitution) without metabolic activation. Freeze-dried samples of *uMakhonya*® were reconstituted in deionised water to known concentrations (5000 μg/mL, 500 μg/mL and 50 μg/mL) prior to biological activity testing. The test was carried out using the plate incorporation procedure described by Maron and Ames[17]. Briefly, 100 μL of bacterial stock was incubated in 20 mL of Oxoid Nutrient Broth for 16 h at 37 °C on an orbital shaker. The overnight culture (100 μL) was added to 2 mL of top agar (containing traces of biotin and histidine) together with 100 μL of test solution (*uMakhonya*® doses, solvent control or positive control) and 500 μL of PBS (for exposure without metabolic activation). The top agar mixture was poured over the surface of the agar plate and incubated for 48 h at 37 °C. After incubation, the number of revertant colonies (mutants) was counted. All cultures were made in triplicate (except the solvent control where up to five replicates were made) for each assay. The positive control used was 4-nitroquinoline-1-oxide (4-NQO) at a concentration of 2 μg/mL. Substances are considered mutagenic if the number of induced revertant colonies is twice the revertant colonies of the negative control (blank).[17]

Cell culture

Normal human whole blood was carefully layered onto equal amounts of Histopaque 1077 then centrifuged at 600 *g* for 30 min at 25 °C. After centrifugation, the buffy coat layer containing PBMCs was isolated and washed twice in PBS (5 mL) and centrifuged again (300 *g* for 20 min at 25 °C). The final pellets were re-suspended in complete culture media (CCM) at 1x10^6 cells/mL and then left untreated or incubated for 2 h with 20 μg/mL of cyclosporine A. Without removing the immunosuppressive effect of cyclosporine, the cells were aliquoted to 6 well plates and treated with doses of *uMakhonya*® ranging from 1000 μg/mL to 10 μg/mL at a ratio of 1:1. The treated PBMCs were then incubated for 24 h at 37 °C, 5% CO_2 and 95% humidity. At the end of the incubation period the cells and their supernatants were used for further experiments.

THP-1 monocytes were cultured in RPMI-1640 with L-glutamine, 10% FCS and 1% PSF in an incubator set at 37 °C with 5% CO_2 and 95% humidity. The cells were passaged every 2–3 days and new media was added. To generate adherent macrophage-like cells, THP-1 cells at a density of 1×10^6 cells/mL were treated with 0.5 μg/mL phorbol 12-myristate 13-acetate (PMA) for 24 h and rested for a further 24 h before stimulation. Confluent wells of cells (1x10^5 cells/mL) were left untreated or activated with 10 μg/mL of LPS from *S. typhosa* for 2 h. Without removing the LPS stimulation, the THP-1 cells were aliquoted to 6 well plates and treated with doses of *uMakhonya*® ranging from 1000 μg/mL to 10 μg/mL at a ratio of 1:1. Cell viability was evaluated after 24 h using the luminescent cell viability ATP assay kit from Promega (USA) while the supernatants of treated cells were used for the nitric oxide free radical scavenging assay. Cyclosporine A was used a positive control for cytotoxicity and nitric oxide secretion suppression.

Cell viability assay

The luminescent cell viability ATP assay kit from Promega uses recombinant luciferase to catalyse the following reaction:

ATP + d-Luciferan + O_2 → Oxyluciferan + AMP + PPi + CO_2 + Light (560 nm).

When ATP is the limiting component in the reaction, the intensity of the emitted light is proportional to the concentration of ATP. Based on these principles, the levels of ATP in cyclosporine immunosuppressed PBMCs, unstimulated, and LPS-stimulated THP-1 cells treated with doses of *uMakhonya*® were analysed according to the manufacturer's instructions. Briefly, a sample (100 μL) of 24 h treated/control cell suspension was pipetted into three different wells of a white opaque 96-well plate. The working CellTiter-Glo™ Reagent (cat number: G7570) was prepared immediately before use and was added to the wells with treated cells at 100 μL per well. The plate was shaken on a plate shaker for 2 min at 150 *g*. This plate was then incubated in darkness for 10 min at room temperature. Background signals of cell culture media and *uMakhonya*® doses (negative control) were subtracted from each average read. A dose response curve was also generated for the ATP levels using RLU versus different concentrations of samples. The cell viability assay was performed in triplicate and repeated three times before the follow up assays were undertaken.

Nitric oxide free radical scavenging activity

The Griess reagent system from Promega (USA) was used to measure nitrite (NO_2^-), which is one of two primary, stable and nonvolatile breakdown products of NO. To perform the assay, 50 μL of separate supernatants from peptidoglycan stimulated, cyclosporine treated (20 μg/mL) and control PBMCs were plated in triplicate 96 well plates. The samples were left to equilibrate to room temperature after which 50 μL of Sulfanilamide solution was dispensed to all experimental samples and incubated for 10 min at room temperature away from light. The NED solution (50 μL) was then dispensed to all sample wells and the plate incubated for another 10 min at room temperature protected from light again. Absorbance was then measured within 30 min in a Zenyth200 plate reader at 540 nm. Nitrite standards at doses ranging from 100 μM to 1.56 μM were included as part of the samples and were used to draw a reference curve. All samples and standards were prepared in triplicate and the experiments were repeated twice.

Malondialdehyde quantification

To measure lipid peroxide levels in treated PBMCs and supernatants, the thiobarbituric acid reactive substances (TBARS) assay was used. Firstly, the treated and control PBMCs in PBS were homogenised on ice and supernatants were centrifuged at 10 000 *g* for 5 min to remove insoluble particles. Cyclosporine A, at 20 μg/mL, was used as a positive control for lipid peroxidation. MDA standards (100 μL of each sample) at doses ranging from 125 μM to 7.8 μM were added into separate microcentrifuge tubes followed by 100 μL of sodium dodecyl sulphate lysis solution. The tubes were mixed thoroughly and incubated for 5 min at room temperature. The TBA reagent (250 μL) was added to each sample and standard and the tubes were incubated at 95 °C for 1 h. After this, the tubes were cooled in an ice bath and then centrifuged at 700 *g* for 15 min. The samples and standards supernatants (200 μL per well of a 96 well plate) were analysed using a Zenyth200 spectrophotometer at 532nm. All samples and standards were read in triplicate and each experiment was repeated twice. A blank control was included to substrate background noise.

1,1-Diphenyl-2-picryl-hydrazyl free radical scavenging activity

The DPPH assay was performed according to Sharma and Bhat[18] with a few deviations. Non-cytotoxic doses of *uMakhonya*® ranging from 10 μg/mL to 100 μg/mL as determined by the ATP cell viability assay on both PBMCs and THP-1 cells were used for this assay. Briefly, samples (3000 μL of different non-cytotoxic doses of *uMakhonya*®, PBS or positive control) and methanolic DPPH solution (1000 μL, 200 μM) were combined and kept in the dark at 37 °C for 30 min. The absorbance of samples was measured at 517 nm on a Zenyth200 plate reader. All tests were performed in triplicate. Ascorbic acid was used as a positive control and was reconstituted in PBS at a concentration of 1 mM (175 μg/mL).

Statistical analysis

Data analysis was done in Microsoft Excel to obtain descriptive statistics. The different levels of significance within the separate treated groups were analysed using one-way analysis of variance (*ANOVA*) and the differences between the treated cells, the untreated cells and the negative control samples were analysed using GraphPad Prism (version 5) software with the Tukey-Kramer multiple comparison test. Differences of $p \leq 0.05$ were considered statistically significant.

Results

Ames test

Results obtained from the mutagenicity test of *uMakhonya*® using *S. typhimurium* TA98 and TA100 strains were expressed as a mean ± s.e.m. (Table 1) and are based on a number of induced revertant colonies. Based on Table 1, *uMakhonya*® doses of up to 5000 μg/mL were not mutagenic in the *Salmonella*/microsome tester strains TA98 and TA100 when compared to the negative control and 4-NQO. The positive control, 4-NQO, induced significantly ($p < 0.05$) higher numbers of colonies and therefore demonstrated its mutagenic activities.

Table 1: Number of histidine positive colonies (revertants) (mean ± s.e.m.) in *Salmonella typhimurium* strains TA98 and TA100 produced by *uMakhonya*®. 4-nitroquinoline-1-oxide (4-NQO) was used as a positive control and deionised water was used as a solvent control. In both test systems *uMakhonya*® was shown to be non-mutagenic.

Test sample name	Concentration (μg/mL)	TA98	TA100
UMakhonya®	5000	15.5±8.0	88.8±10.9
	500	14.6±2.3	87.4±12.5
	50	15.6±4.1	90.5±7.6
4 NQO	2	373.8±59.4	893.7±100.1
Solvent		13.0±3.0	97.6±8.9

Cell viability assay

The cytotoxicity of *uMakhonya*® was evaluated in normal human PBMCs and malignant THP-1 monocytes to establish noncytotoxic doses for the antioxidant assays. In PBMCs immunosuppressed with 20 μg/mL of cyclosporine, *uMakhonya*® induced a dose dependent cytotoxic effect with high doses (1000 μg/mL and 500 μg/mL) significantly ($p<0.05$) increasing immunosuppression. Lower doses were less cytotoxic and these were used for the evaluation of MDA levels (Figure 1).

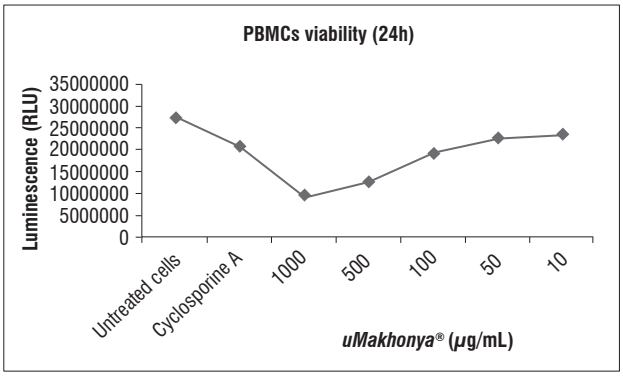


Figure 1: Cytotoxic effects of *uMakhonya*® doses on isolated PBMCs immunosuppressed with 20 μg/mL of cyclosporine A.

THP-1 monocytes stimulated with LPS from *S. typhi* and then treated with doses of *uMakhonya*® showed a similar trend to PBMCs. Higher doses were significantly ($p < 0.05$) cytotoxic to stimulated and unstimulated monocytes.

Table 2: Evaluation of nitrite as a measure of nitric oxide secretion from LPS stimulated THP-1 monocytes.

Treatment type	Untreated THP-1 cells	LPS (10 μg/mL) stimulated THP-1 cells	Cyclosporine A (20 μg/mL)	*uMakhonya*® (100 μg/mL)	*uMakhonya*® (50 μg/mL)	*uMakhonya*® (10 μg/mL)
Nitrite (μM) (mean ± s.e.m.)	1.03± 0.0045	2.12± 0.0069	2.00± 0.0068	1.63± 0.0025	2.24± 0.0023	2.12±0.00

At these higher doses *uMakhonya*® was more cytotoxic than the immunosuppressive drug cyclosporine (Figure 2). Supernatants from stimulated monocytes treated with lower doses were used for the evaluation of nitrite radicals as a measure of the radical scavenging potential of *uMakhonya*®.

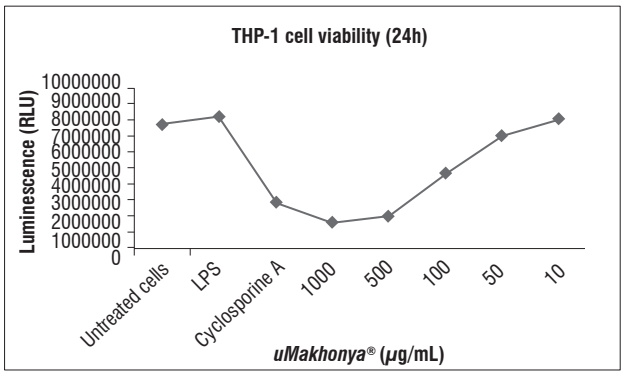


Figure 2: Cell viability of LPS stimulated THP-1 monocytes treated with doses of *uMakhonya*® ranging from 1000 μg/mL to 10 μg/mL.

Nitric oxide free radical scavenging activity

Nitrite (NO_2^-) is one of two primary, stable and non-volatile breakdown products of NO. Stimulation of THP-1 monocytes with LPS significantly ($p<0.05$) increased the secretion of NO as represented by levels of NO_2^-. Treatment of LPS stimulated THP-1 monocytes with non-cytotoxic doses of *uMakhonya*® did not significantly ($p> 0.05$) change the levels of nitrite radicals in treated supernatants with the highest dose tested (100 μg/mL) showing potential radical scavenging potential but this was not significantly ($p>0.05$) different from LPS stimulated THP-1 monocytes (Table 2).

Malondialdehyde quantification

Immunosuppression of PBMCs with cyclosporine was meant to increase lipid peroxidation and changes in levels of peroxides after treatment with *uMakhonya*® were a measure of the possible antioxidant effects of this traditional medicinal product. In PBMCs, treatment with *uMakhonya*® doses increased lipid peroxides and this was significant ($p<0.05$) at the lowest dose (10 μg/mL) when compared to PBMCs treated with cyclosporine only. In supernatants, *uMakhonya*® doses significantly ($p<0.05$) decreased lipid peroxides at all doses tested when compared to supernatants from PBMCs treated with cyclosporine A only. Lipid peroxide levels from supernatants of immunosuppressed PBMCs were similar to those of untreated supernatants (Figure 3). Therefore *uMakhonya*® increased intracellular levels of lipid peroxides while displaying potent abilities to scavenge for these radicals in the surrounding media of the treated PBMCs.

1,1-Diphenyl-2-picryl-hydrazyl free radical scavenging activity

The DPPH assay is one of the quick methods used to evaluate antioxidant activity on DPPH, a stable free radical and widely used index. DPPH salt at 200 μM in methanol was mixed with non-cytotoxic doses of *uMakhonya*® in PBS ranging from 100 to 10 μg/mL and the changes in absorbance were measured as an indicator of free radical scavenging activity. Ascorbic acid was reconstituted in the same medium as *uMakhonya*® (PBS) to ensure that the results of the free radical scavenging activities were comparable.

Ascorbic acid had a reduced antioxidant effect when dissolved in PBS with a plateau being reached at 80 μM (14 μg/mL) at 25% reduction of DPPH. Therefore, the IC_{50} for free radical scavenging activity for ascorbic acid was not reached when PBS was used as reconstitution medium (Figure 4a). There was a dose dependent increase in DPPH reduction by *uMakhonya*® with the highest dose tested (100 μg/mL) showing significantly ($p<0.05$) better antioxidant potential than ascorbic acid at 14 μg/mL. Similarly to ascorbic acid, *uMakhonya*® also did not reach an IC_{50} dose when doses shown to be noncytotoxic to PBMCs and THP-1 monocytes used were tested (Figure 4b).

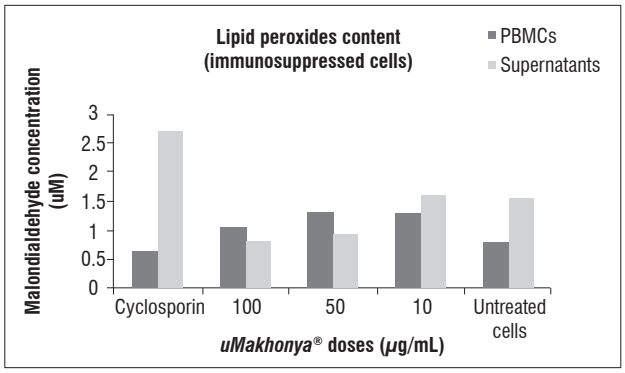


Figure 3: Effects of noncytotoxic doses of *uMakhonya*® on levels of lipid peroxides in cyclosporine treated peripheral blood mononuclear cells (PBMCs) and their supernatants.

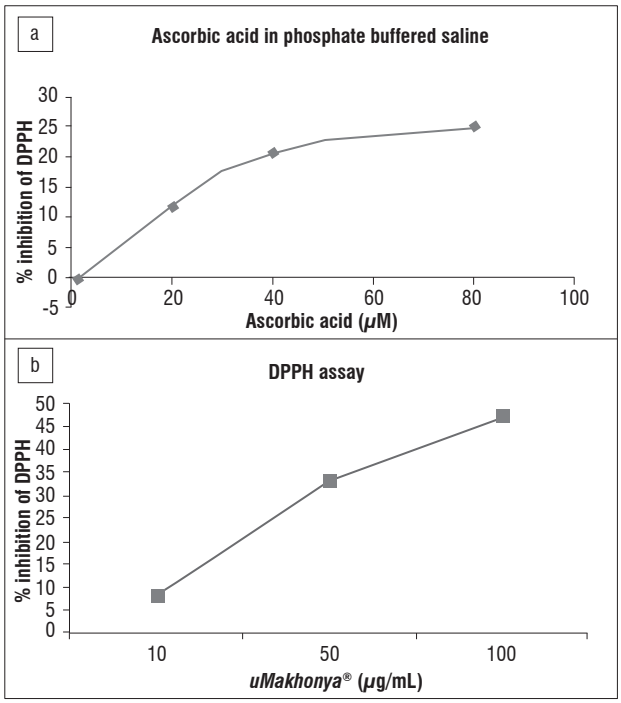


Figure 4: Reduction of 1,1-diphenyl-2-picryl-hydrazyl (DPPH) salt (200 μM) as a measure of (a) the antioxidant potential of ascorbic acid doses (b) and *uMakhonya*® doses

Discussion

We tested the possible mutagenic and antioxidant effects of *uMakhonya*®, one of the popular commercial immune boosters based on traditional medicine knowledge. Even though it is a difficult task to convince manufacturers or herbalists to reveal the ingredients as well as the recipes of their products as they are well guarded secrets,[7] this study collaborated with the owner of the traditional medicinal product. Novel potential drugs are usually screened for their possible mutagenic activities in many systems, including the *Salmonella typhimurium* microsome assay.[19] Using the Ames test, *uMakhonya*® doses of up to 5000 μg/mL were not mutagenic in the *Salmonella*/microsome tester strains TA98 and TA100 when compared to the solvent negative control and 4-NQO (Table 1). These results showed that *uMakhonya*® can be assumed to have no direct in-vitro mutagenic effects when not metabolically activated. Similar results were shown in a study of commercial traditional herbal products sold in KwaZulu Natal where the Ames test results revealed that all 14 herbal preparations were non-mutagenic towards the *Salmonella typhimurium* strains TA98 and TA100 without metabolic activation.[20] Mutagenic effects of any extract can be manifested in two other ways: the extract can be promutagenic by requiring metabolic activation or comutagenic by enhancing the mutagenic response of a known mutagen, irrespective of being mutagenic or not.[21] Therefore, the risk assessment of *uMakhonya*® is not complete until this traditional medicinal product is tested for promutagenic and comutagenic effects.

Cell defences against oxidative stress are also known to decrease through changes in gene expression. The greatest protection against mutagenic activity observed in the promutagenic assay may be related to the activation of cytochrome P450 which mediates the oxidation of promutagens. Cytochromes function as antioxidants which scavenge and neutralise compounds that generate oxygen radicals, free radicals and reactive oxygen species.[9] Most medicinal plants are known to possess innate antioxidant activities without the need for metabolic activation. Antioxidants from dietary and medicinal plant sources, particularly those containing phenolic compounds, have significant antioxidant activity.[6] Mutagenic and antimutagenic activities have been linked to the presence of certain phytochemical substances such as flavonoids and tannins.[9] The lack of direct cytotoxicity of *uMakhonya*® at doses of up to 5000 μg/mL observed in the Ames test was contrasted with direct cytotoxicity on isolated immune cells to establish non-cytotoxic doses for radical scavenging activities. In immunosuppressed PBMCs, *uMakhonya*® induced a dose-dependent cytotoxic effect at high doses above 100 μg/mL thus increasing immunosuppression significantly (Figure 1). A similar trend was observed in LPS stimulated THP-1 macrophages with similarly high doses as cytotoxic as the immunosuppressive drug cyclosporine A (Figure 2). *UMakhonya*® doses ranging from 10 μg/mL to 100 μg/mL were chosen to assess the antioxidant activities of this traditional medicine product.

Engagement of cellular receptors by LPS in stimulated immune cells leads to synthesis of new proteins through alteration in the pattern of gene expression.[22] NO synthesis and secretion increases considerably after exposure to immunological stimuli such as bacterial LPS.[23] Secretion of NO was induced by stimulation of THP-1 monocytes with LPS from *S. typhi* followed by treatment with doses of *uMakhonya*® for 24 hours. The Griess reagent system measures nitrite radicals as a reflection of the amount of NO secreted by cells. Treatment with non-cytotoxic doses of *uMakhonya*® showed a dose dependent effect on the levels of nitrite radicals in treated supernatants with the highest dose tested (100 μg/mL) showing possible radical scavenging potential by significantly ($p<0.05$) decreasing nitrite radicals when compared to LPS stimulated THP-1 monocytes. NO is a physiological mediator produced by many cells involved in immunity and inflammation. When generated in high concentrations, NO is rapidly oxidised to reactive nitrogen oxide species (RNOS) that mediate most of the immunological effects of NO. RNOS can reduce thiols to modify key signalling molecules such as kinases and transcription factors.[24] Cyclosporine A, a known inhibitor of induced nitric oxide synthase, did not cause a significant ($p>0.05$) decrease in nitrite radicals when compared to LPS stimulated THP-1 cells. This may be related to the dose of cyclosporine A used which in this case was not effective in suppressing NO secretion. From the observed results, *uMakhonya*® did not show potent antioxidant activity, with only the highest non-cytotoxic dose showing nitrite radical scavenging potential.

The antioxidant activity of a given extract depends not only on its chemical constituents but also on the type of generated radical it can neutralise.[6] For this reason we also tested the radical scavenging activity of *uMakhonya*® using the TBARS and DPPH assays. The occurrence of lipid peroxidation in biological membranes causes impairment of membrane functioning, changes in fluidity, inactivation of membrane-bound receptors and enzymes, and increased non-specific permeability to ions such as calcium (Ca^{2+}).[25] Lipid peroxidation was induced by treating PBMCs with cyclosporine A followed by incubation with various non-cytotoxic doses of *uMakhonya*®. This traditional medicinal product showed significant antioxidant potential by reducing lipid peroxides in supernatants of immunosuppressed cells to levels of the untreated control supernatants. Such antioxidant activity was not seen inside the immunosuppressed PBMCs treated with several doses of *uMakhonya*®. In the DPPH assay *uMakhonya*® showed a dose dependent increase in radical scavenging activity with the highest dose (100 μg/mL) showing the greatest potential (Figure 4b). The radical scavenging profile of ascorbic acid reached a plateau at 80 μM (14 μg/mL) without reaching an IC_{50} concentration when this known antioxidant was reconstituted in PBS (Figure 4a), the same medium used to reconstitute freeze dried samples of *uMakhonya*®. Other studies have shown that ascorbic acid has IC_{50} values of 11.8 μM[18] and 56 μM[26] when dissolved in methanol. Therefore, the use of an aqueous solution like PBS to dissolve both ascorbic acid and *uMakhonya*® might limit their radical scavenging activities.

Lipid peroxides are one of the aldehyde products of reactive oxygen species degradation of membrane lipids and these aldehydes can cause cross links in nucleic acids leading to DNA damage.[27] The results of this study showed that *uMakhonya*® did not induce significant intracellular lipid peroxidation while showing significant radical scavenging activities in the supernatants of treated immunosuppressed PBMCs and in the DPPH assay. A recent study of a South African traditional herbal product with similar claims to *uMakhonya*® but of unknown composition, showed promising antioxidant potential in the DPPH assay, ferric reducing power and β-carotene-linoleic acid model system.[7] Different studies have shown that the individual medicinal plants constituting *uMakhonya*® possess both antioxidant and antimutagenic activities. Dichloromethane and 90% methanol extracts of *A. afra* did not show mutagenicity in strain TA98 with and without metabolic activation.[28] Volatile oils from *A. afra* have also been shown to possess considerable antioxidant effects in preventing the discoloration of β-carotene and linoleic acid and also showed significant radical scavenging potential in the lipid peroxidation assay.[29,30] *C. crispus* displayed antioxidant activities in the DPPH and ferric-reducing antioxidant power (FRAP) assays and contained phenolics, condensed tannins and flavonoids.[30,31] *C. crispus* extracts have not been shown to possess any genotoxic effects.[33] *P. guajava* leaves and bark extracts showed concentration-dependent scavenging activity on hydrogen peroxide, superoxide and DPPH.[34] Pre-treatment with an aqueous guava leaf extract was found to be effective in inactivating the mutagenicity of direct-acting mutagens 4-nitro-o-phenylenediamine and 2-aminofluorene in the tester strains of *Salmonella typhimurium.*[35] Freeze-dried extracts of *U. tomentosa* have been shown to have significant antioxidant activities in scavenging free radicals in the DPPH and ABTS assays.[36] *U. tomentosa* extracts were able to inhibit 90% of the mutagenic effect of hydrogen peroxide and did not show significant genotoxicity.[37] *U. tomentosa* showed significant antioxidant activities in the trolox equivalent antioxidant capacity (TEAC), peroxyl radical-trapping capacity (PRTC), and superoxide radical scavenging activity (SOD) assays and these were attributed to the total phenolics and tannins content of the extract.[38] Menthol has been shown to have no genotoxic effects in both in-vitro and in-vivo animal models.[39] *Mentha pepirita* extract which contains menthol has also been shown to have antioxidant and antiperoxidant properties.[40] Although all these herbal plants have been proven to have antimutagenic and antioxidant effects, it is not easy to predict their behaviour when combined together.

Herbalists have known for centuries the value of using a combination of herbal remedies, single extracts and combined extracts to switch on the body's defence mechanisms, self-healing and protective processes.[1]

Conclusion

This was the first study to look at the mutagenic and antioxidant effects of the immune booster, *uMakhonya®,* which is prepared by combining five different traditional medicinal plants. The hot water extract of *uMakhonya®* did not show significant mutagenicity in the *Salmonella*/microsome tester strains TA98 and TA100 when compared to the solvent negative control and 4-NQO positive control. In immunosuppressed PBMCs supernatants *uMakhonya®* showed significant radical scavenging activity in reducing lipid peroxides but this ability was reduced in the PBMCs. *UMakhonya®* showed a dose dependent reduction in nitrite radicals in LPS stimulated THP-1 monocytes. In the DPPH assay, *uMakhonya®* showed better radical scavenging activity than ascorbic acid when both were re-dissolved in aqueous PBS. Therefore *uMakhonya®* extract was not mutagenic and showed promising antioxidant activity in immune cells. Future studies should focus on the mutagenic effects of this product during metabolic activation and in-vivo antioxidant effects.

Acknowledgements

We would like to thank the National Research Foundation for providing financial assistance for this project, the South African National Blood Service (SANBS) for the kind donation of normal human whole blood reagents and Mr Saiyur Ramsugit from the discipline of Medical Microbiology at the University of KwaZulu Natal for the donation of the THP-1 monocytes. We would also like to thank Dr Esam Elgorashi of the Agricultural Research Council for his assistance with the Ames test.

Authors' contributions

M.N. conducted all the experiments and drafted the primary manuscript. N.G. was principal investigator and conceptualised and supervised the research, and contributed in the final writing of the manuscript. V.N. was co-investigator and contributed to the analysis and the writing of the manuscript. V.N. has a wealth of knowledge on the formulation of the product and its clinical effects on consumers.

References

1. Busia K. Medical provision in Africa – past and present. Phytother Res. 2005;19(11):919–923. http://dx.doi.org/10.1002/ptr.1775
2. Olivier DK. The ethnobotany and chemistry of South African tonic plants [PhD dissertation]. Johannesburg: University of Johannesburg; 2012.
3. Okpako DT. Traditional African medicine: Theory and pharmacology explored. Trends Pharmacol Sci. 1999;20(12):482–485. http://dx.doi.org/10.1016/S0165-6147(99)01406-6
4. Chen X, Howard OMZ, Yang X, Wang L, Oppenheim JO, Krakauer T. Effects of *Shuanghuanglian* and *Qingkialing*, two multi-components of traditional Chinese medicinal preparations, on human leukocyte function. Life Sci. 2002;70(24):2897–2913. http://dx.doi.org/10.1016/S0024-3205(02)01541-2
5. Tshibangu KC, Worku ZB, De Jongh MA, Van Wyk AE, Mokwena SO, Peronovic V. Assessment of effectiveness of traditional medicine in managing HIV/AIDS patients in South Africa. East Afr Med J. 2004;81(10):499–504. http://dx.doi.org/10.4314/eamj.v81i10.9231
6. Boubaker J, Mansour HB, Ghedira K, Chekir-Ghedira L. Antimutagenic and free radical scavenger effects of leaf extracts from *Accacia salicina*. Ann Clin Microbiol Antimicrob. 2011;10(37):1–10. http://dx.doi.org/10.1186/1476-0711-10-37
7. Ndhlala AR, Finnie JF, Van Staden J. In-vitro antioxidant properties, HIV-1 reverse transcriptase and acetylcholinesterase inhibitory effects of traditional herbal preparations sold in South Africa. Molecules. 2010;15(10):6888–6904. http://dx.doi.org/10.3390/molecules15106888
8. Cowan MM. Plant products as antimicrobial agents. Clin Microbiol Rev. 1999;12(4):564–582.
9. Horn RC, Vargas VMF. Antimutagenic activity of extracts of natural substances in the *Salmonella*/microsome assay. Mutagenesis. 2003;18(2):113–118. http://dx.doi.org/10.1093/mutage/18.2.113
10. Ferguson LR. Antimutagens as cancer chemopreventive agents in diet. Mutat Res. 1994;307(1):395–410. http://dx.doi.org/10.1016/0027-5107(94)90313-1
11. Lakshmi B, Ajith TA, Jose N, Janardhanan KK. Antimutagenic activity of methanolic extract of *Ganode lucidum* and its effects on hepatic damage caused by benzo[a]pyrene. J Ethnopharmacol. 2006;107(2):297–303. http://dx.doi.org/10.1016/j.jep.2006.03.027
12. Eccles R. Menthol and related cooling compounds. J Pharm Pharmacol. 1994;46(8):618–630. http://dx.doi.org/10.1111/j.2042-7158.1994.tb03871.x
13. Holetz FB, Pessini GL, Sanches NR, Cortez DAG, Nakamura CV, Dias Filho BP. Screening of some plants used in the Brazilian folk medicine for the treatment of infectious diseases. Mem Inst Oswaldo Cruz. 2002;97(7):1027–1031. http://dx.doi.org/10.1590/S0074-02762002000700017
14. Liu D, Keesing JK, He P, Wang Z, Shi Y, Wang Y. The world's largest macroalgal bloom in the Yellow Sea, China: Formation and implications. Estuar Coast Shelf Sci. 2013;129(1):2–10. http://dx.doi.org/10.1016/j.ecss.2013.05.021
15. Okai Y, Higashi-Okai K, Ishizaka S, Yamashita U. Enhancing effect of polysaccharides from an edible brown alga, *Hijikia fusiforme* (Hijiki), on release of tumor necrosis factor-α from macrophages of endotoxin-nonresponder C3H/Hej mice. Cancer Lett. 1997;27(1):74–79. http://dx.doi.org/10.1080/01635589709514505
16. Bors M, Michalowicz J, Pilarski R, Sicińska P, Gulewicz K, Bukowska B. Studies of biological properties of *Uncaria tomentosa* extracts on human blood mononuclear cells. J Ethnopharmacol. 2012;142(3):669–678. http://dx.doi.org/10.1016/j.jep.2012.05.036
17. Maron DM, Ames BN. Revised methods for the *Salmonella* mutagenicity test. Mutat Res. 1983;113(3–4):173–215. http://dx.doi.org/10.1016/0165-1161(83)90010-9
18. Sharma OP, Bhat TK. DPPH antioxidant assay revisited. Food Chem. 2009;113(4):1202–1205. http://dx.doi.org/10.1016/j.foodchem.2008.08.008
19. El-Sayed WM, Hussin WA. Antimutagenic and antioxidant activity of novel 4-substituted phenyl-2,2'-bichalcophenes and aza-analogs. Drug Des Devel Ther. 2013;7(1):73–81. http://dx.doi.org/10.2147/DDDT.S40129
20. Ndhlala AR. Pharmacological, phytochemical and safety evaluation of commercial herbal preparations common in South Africa [PhD dissertation]. Pietermaritzburg: University of KwaZulu Natal; 2009.
21. Snijman PW, Swanevelder S, Joubert E, Green IR, Gelderblom WCA. The antimutagenic activity of the major flavonoids of rooibos (*Aspalathus linearis*): Some dose-response effects on mutagen activation-flavonoids interactions. Mutat Res. 2007;631(2):111–123. http://dx.doi.org/10.1016/j.mrgentox.2007.03.009
22. Freudenberg MA, Tchaptchet S, Keck S, Fejer G, Huber M, Schütze N, et al. Lipopolysaccharide sensing an important factor in the innate immune response to Gram-negative bacterial infections: Benefits and hazards of LPS hypersensitivity. Immunobiology. 2008;213(3–4):193–203. http://dx.doi.org/10.1016/j.imbio.2007.11.008
23. Khatsenko OG, Gross SS, Rifkind AB, Vane JR. Nitric oxide is a mediator of the decrease in cytochrome P450-dependent metabolism caused by immunostimulants. P Natl Acad Sci USA. 1993;90(23):11147–11151. http://dx.doi.org/10.1073/pnas.90.23.11147
24. Coleman JW. Nitric oxide in immunity and inflammation. Int Immunopharmacol. 2001;1(8):1397–1406. http://dx.doi.org/10.1016/S1567-5769(01)00086-8
25. Niki E, Yoshida Y, Saito Y, Noguchi N. Lipid peroxidation: Mechanisms, inhibition, and biological effects. Biochem Biophys Res Commun. 2005;338(1):668–676. http://dx.doi.org/10.1016/j.bbrc.2005.08.072
26. Kano M, Takayanagi T, Harada K, Makino K, Ishikawa F. Antioxidative activity of anthocyanins from purple sweet potato *Ipomoera batatas* cultivar Ayamurasaki. Biosci Biotechnol Biochem. 2005;69(5):979–988. http://dx.doi.org/10.1271/bbb.69.979
27. Tuzgen S, Hamnimoglu H, Tanriverdi T, Kacira T, Sanus G, Atukereny P, et al. Relationship between DNA damage and total antioxidant capacity in patients with glioblastoma multiforme. Clin Oncol. 2007;19(3):177–181. http://dx.doi.org/10.1016/j.clon.2006.11.012

28. Elgorashi EE, Taylor JLS, Maes A, Van Staden J, De Kimpe N, Verschaeve L. Screening of medicinal plants used in South African traditional medicine for genotoxic effects. Toxicol Lett. 2003;143(2):195–207. http://dx.doi.org/10.1016/S0378-4274(03)00176-0

29. Patil GV, Dass SK, Chandra R. *Artemisia afra* and modern diseases. Pharmacogenom Pharmacoproteomics. 2011;2(3):1–22. http://dx.doi.org/10.4172/2153-0645.1000105

30. Burits M, Asres K, Bucar F. The antioxidant activity of the essential oils of *Artemisia afra*, *Artemisia abyssinica* and *Juniperus procera*. Phytother Res. 2001;15(2):103–108. http://dx.doi.org/10.1002/ptr.691

31. Jimenez-Escrig A, Jimenez-Jimenez I, Pulido R, Saura-Calixto F. Antioxidant activity of fresh and processed edible seaweeds. J Sci Food Agr. 2001;81(5):530–534. http://dx.doi.org/10.1002/jsfa.842

32. Cox S, Abu-Ghannam N, Gupta S. An assessment of the antioxidant and antimicrobial activity of six species of edible Irish seaweeds. Intern Food Res J. 2010;17(1):205–220.

33. World Health Organization International Programme on Chemical Safety. Safety evaluation of certain food additives: Carrageenan. Geneva: WHO Food Additives Series:60; 1999. p. 1–50.

34. Ogunlana, OE, Ogunlana OO. In-vitro assessment of the free radical scavenging activity of *Psidium Guajava*. Res J Agric Bio Sci. 2008;4(6):666–671.

35. Gutierrez RMP, Mitchell S, Solis RV. *Psidium guajava*: A review of its traditional uses, phytochemistry and pharmacology. J Ethnopharmacol. 2008;117(1):1–27. http://dx.doi.org/10.1016/j.jep.2008.01.025

36. Sandoval M, Okuhama NN, Zhang XJ, Condezo LA, Lao J, Angeles FM, et al. Anti-inflammatory and antioxidant activities of cat's claw (*Uncaria tomentosa* and *Uncaria guianensis*) are independent of their alkaloid content. Phytomedicine. 2002;9(4):325–337. http://dx.doi.org/10.1078/0944-7113-00117

37. Romero-Jimenez M, Campos-Sanchez J, Analla M, Munoz-Serrano A, Alonso-Moraga A. Genotoxicity and anti-genotoxicity of some traditional medicinal herbs. Mutat Res. 2005;585(1–2):147–155. http://dx.doi.org/10.1016/j.mrgentox.2005.05.004

38. Pilarski R, Zielinski H, Ciesiołka D, Gulewicz K. Antioxidant activity of ethanolic and aqueous extracts of *Uncaria tomentosa* (Willd.) DC. J Ethnopharmacol. 2005;104(1–2):18–23. http://dx.doi.org/10.1016/j.jep.2005.08.046

39. World Health Organization International Programme on Chemical Safety. Safety evaluation of certain food additives: Menthol. Geneva: WHO Food Additives Series: 60; 1999. p. 1–47.

40. Maraj M, Ali N, Rashid U. Antimutagenic properties of *Mentha piperita* extract against ethyl methane sulphonate induced mutagenicity in *Mus musculus*. Int J Pharm Sci Rev Res. 2010;5(2):63–66.

Trading on extinction: An open-access deterrence model for the South African abalone fishery

AUTHOR:
Douglas J. Crookes[1]

AFFILIATION:
[1]Department of Economics, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Douglas Crookes

EMAIL:
dcrookes@outlook.com

POSTAL ADDRESS:
Department of Economics, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

DATES:
Received: 23 June 2015
Revised: 05 Aug. 2015
Accepted: 09 Oct. 2015

KEYWORDS:
rhino; system dynamics; illegal harvesting; predator–prey; Table Mountain National Park

HOW TO CITE:
Crookes DJ. Trading on extinction: An open-access deterrence model for the South African abalone fishery. S Afr J Sci. 2016;112(3/4), Art. #2015-0237, 9 pages. http://dx.doi.org/10.17159/sajs.2016/20150237

South African rhinoceros (e.g. *Diceros bicornis*) and abalone (*Haliotis midae*) have in common that they both are harvested under open-access conditions, are high-value commodities and are traded illegally. The difference is that a legal market for abalone already exists. An open-access deterrence model was developed for South African abalone, using Table Mountain National Park as a case study. It was found that illegal poaching spiked following the closure of the recreational fishery. The resource custodian's objective is to maximise returns from confiscations. This study showed that a legal trade results in a 'trading on extinction' resource trap, with a race for profits, an increase in the probability of detection *after* a poaching event and the depletion of populations. In contrast with HS Gordon's seminal article (J Polit Econ 1954;62:124–142), profit maximisation does not automatically improve the sustainability of the resource. Under certain conditions (e.g. a legal trade with costly enforcement), profit maximisation may actually deplete abalone populations. The article also has implications for rhino populations, as a legal trade is currently proposed.

Introduction

The South African abalone (*Haliotis midae*) fishery is experiencing a crisis. Illegal harvesting of abalone has escalated dramatically in recent years, to such an extent that the fishery was closed between 2008 and 2010.[1] Several reasons for this escalation have been presented. For example, Raemaekers et al.[2] argue that both the rise in abalone prices in the 1990s and the failure to include traditional fishers in the reform process, were drivers for the collapse in the fishery. Muchapondwa et al.[1], however, found that drug use and corruption were major factors in the illegal trade in abalone.

The abalone fishery has a number of important features that characterise systems of illegal exploitation. Firstly, there is what is known as open-access to the resource. Here open-access implies that entrants have relatively free access to the resource and that species are subject to low levels of protection and enforcement and are subject to poaching.[3] Secondly, the authorities seek to restrict access to the resource by patrolling the fishery, arresting offenders and enforcing some sort of penalty for the offence. Thirdly, the fishery is characterised by a complex system.[2] This characterisation requires some form of integrated modelling to develop not only the biological dynamics of the abalone population, but also the dynamics of fishing effort.

Therefore, a model is developed here to capture each of these three features. The model is an open-access deterrence model and is developed using system dynamics modelling. Optimisation is used to calibrate the model to fit the historical data, so that unknown biological parameters may be estimated. A unique feature of the model is that it encapsulates decision-making by the resource custodian (in this case, the marine authority), thus enabling one not only to assess the implications for management of the abalone fishery, but also to see the potential implications for rhino management, as both rhino management and abalone management contain a number of similar features that will be discussed. The model is then used to answer 'what if' type policy questions to improve the sustainability of the abalone population.

Study site

The model is developed for the Table Mountain National Park in the Western Cape, South Africa, which falls within Zone E (Figure 1). Before the park was proclaimed in 2004, it was known as the Cape Peninsula National Park. In the same year the coastal waters surrounding the park were proclaimed as a marine protected area. However, fishing is still allowed in certain areas. This area was selected because it is close to an urban area (the City of Cape Town), has a good mix of legal fishing and illegal fishing, and is in close proximity to a number of fishing communities. The western areas of South Africa house some of the last remaining viable abalone populations, as many of the abalone populations to the southeast of the study area have already been depleted.

The model

Model attributes

The model comprises four features: (1) open-access modelling of poacher–population interactions using the Gordon–Schaefer model; (2) a deterrence model of poacher behaviour based on Becker's optimal enforcement model; (3) the marine authority's response to poaching; and (4) modelling of the fishery as an integrated complex system.

Open-access modelling

The Gordon–Schaefer open-access model is well described in the literature.[4,5] Changes in stocks are a function of a biological growth f(x) less harvests (h):

$$\dot{x}=f(x)-h \qquad \text{Equation 1}$$

Figure 1: Map of the study site, showing Zone E within the context of other abalone fishing zones in the Western Cape Province of South Africa.

Most frequently, f(x) is the logistic function although other density-dependent variants are possible.[6] The harvests are represented by the standard Schaefer production function, which is a linear function of effort (E) and stock size (x):

$$h(E,x) = qEx. \quad \text{Equation 2}$$

Here q is the catchability coefficient. Effort dynamics under open-access harvesting is a function of profitability, where greater profits attract additional entrants and fewer profits discourage entrants:

$$\dot{E} = \emptyset\pi, \quad \text{Equation 3}$$

where π is a profit function relating prices to costs and is an adjustment parameter determining the rate at which effort adjusts to changes in profitability. In this model, an adapted adjustment parameter $n' = \emptyset p$ is used, where p is the unit price of abalone, expressed as USD/tonnes. Costs on the other hand are a function of the effort expended in order to catch the resource.

Economics of crime and punishment

In a seminal article by Nobel laureate Gary Becker, it is argued that the supply of offences is a function of the probability of conviction per offence, and also the punishment per offence.[7] The model has many applications in the economics literature. One example is an application to the economics of fraud by Van Heerden et al.[8] This theory has been applied to fisheries by Sutinen and Anderson[9], and more recently by Kuperan and Sutinen[10]. The Becker model was also used by Muchapondwa et al.[1] to model the abalone fishery in South Africa.

Marine authority's objective

A unique feature of the model is the way that the marine authority's response to the poaching crisis is modelled. According to Goga[11], up to 30% of the marine authority's operating budget is obtained through the sale of confiscated abalone on the open market. Revenues from the trade in abalone therefore play an important role in enforcement efforts. This role is modelled in the study by adopting a standard neoclassical approach, where the aim of the marine authority is to adapt enforcement effort in order to maximise the revenues obtained through the sale of confiscated abalone, subject to the constraints imposed by the abalone population as well as poaching effort. Some would argue that it is inappropriate for an authority tasked with resource conservation to maximise revenues from confiscations. However, this is very much aligned with the current emphasis of environmental authorities on sustainable resource use initiatives rather than conservation per se. It is therefore necessary to test the feasibility of this profit maximising objective for the sustainability of abalone stocks.

System dynamics modelling

The abalone fishery in South Africa may be viewed as a complex system, with many actors and role players. A system dynamics modelling approach provides a way of capturing those dynamics. A model was constructed to capture the interactions between poachers, the crew members, boat owners and the abalone population (Figure 2). The historical data span from 1980 to 2007, a time period of 27 years. The model is projected forward to the year 2100 to consider the effects of different management strategies on the dynamics of both abalone stocks and fishing effort.

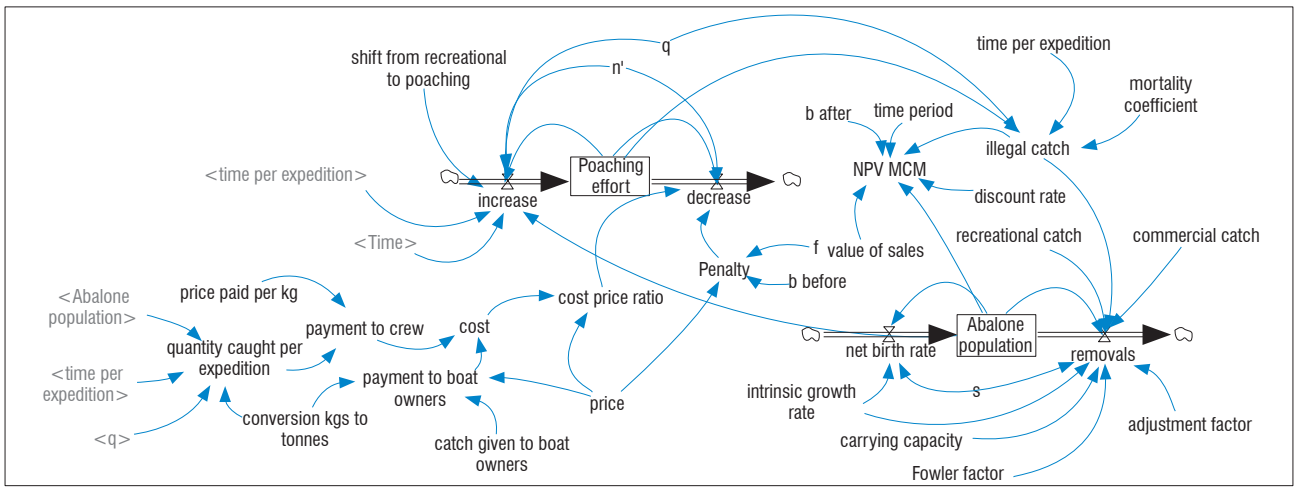


Figure 2: Stock flow diagram for the abalone fishery.

Data

The data were derived from a number of sources. Edwards and Plagányi[12] was the primary source of time series data on commercial catch per unit effort (CPUE) and commercial, recreational and illegal catches from 1980 to 2007. CPUE was standardised using a general linear modelling technique to eliminate differences between catch rates that are a result of any factor besides chance or differences in abundance (for example see Allen and Punsly[13]). These data were supplemented by data on illegal catches from 2003 to 2005[14] which were taken from confiscation records in Table Mountain National Park. In order to derive an estimate for the total illegal catch it was necessary to divide these amounts by the proportion of confiscations (b^D).

It is unlikely that this proportion would be higher than 0.25, as data from Zones A to D suggest a range of between 0.06 and 0.25 (see Brandão and Butterworth[15]). In order to generate a lower bound estimate for illegal catch, the maximum proportion of confiscations was assumed to be 0.25. Therefore, illegal catch for the years 2003 to 2005 is estimated using:

$$\text{Illegal catch}_t = \text{Confiscations}_t / b^D, \text{ for } t = 2003,..,2005, \quad \text{Equation 4}$$

where $b^D = 0.25$ (Table 1). For the 2007 fishing season, it was assumed that trends in other zones were indicative of the situation in Zone E. The illegal catch in Zones A to D amounted to 927 tonnes in 2007[15] and the total allowable catch (TAC) was 125 tonnes[16].

Table 1: Exogenous parameters used in the South African abalone fishery model

Parameter	Symbol	Value	Unit	Reference
Price of abalone for poachers	p	22 000	US dollars/tonne	De Greef and Raemaekers[31]
Time per expedition	T_E	240	Minutes/expedition	De Greef[17]
Biomass in base year (1980)	X_{1980}	1484	Tonnes	Edwards and Plagányi[12]
Poaching effort in base year (1980)	E_{1980}	10	Expeditions	Own calculations
Time frame of historical data for commercial catches	T_C	27	Years	Edwards and Plagányi[12]
Time frame of historical data for recreational catches	T_R	23	Years	Edwards and Plagányi[12]
Shift from recreational to poaching		290	Expeditions/year	Optimisation
Penalty for repeat offender	f	2000	US dollars	De Greef[17]
Price paid to crew and carriers for abalone		5	US dollars/kilogram	Calculation based on De Greef and Raemaekers[31]
Catch given to boat owners	c_B	10	Kilograms/expedition	De Greef and Raemaekers[31]
Commercial catch coefficient		0.0055	Dimensionless	Calculation based on Edwards and Plagányi[12]
Catchability coefficient	q	9.3e-007	1/minutes	Edwards and Plagányi[12]
Proportion of confiscations	b^D	0.25	Dimensionless	Maximum based on Brandão and Butterworth[15]
Probability of detection prior to offence	b	0.025	Dimensionless	Assumption (tested through sensitivity analysis)
Adjustment coefficient	n'	0.44	Expeditions/tonnes/year	Optimisation
Carrying capacity	k	1710	Tonnes	Edwards and Plagányi[12]
Fowler (curvilinear) factor	z	0.88	Dimensionless	Optimisation
Intrinsic growth rate	r	0.1	Dimensionless	Edwards and Plagányi[12]
Discount rate	δ	0.01	1/year	Assumption (tested through sensitivity analysis)
Value of sales	V^A	1800	US dollars/tonnes	Goga[11]

The TAC in Zone E was 12 tonnes in 2007, so apportioning by TAC gives an estimated illegal catch in Zone E of 89 tonnes in 2007 (12/125*927). Table 2 shows the poaching data for 2003–2007, estimated in this way. The methodology is coarse but shows that illegal fishing has increased dramatically over the period since 2003. The anecdotal evidence supports this assertion.[1,2,16]

Illegal fishing effort is estimated by dividing CPUE data[12] into the illegal catch for that year. The CPUE data are for commercial fisheries. In the absence of other available data, it is necessary to assume that the CPUE for the illegal fishery would be similar. CPUE is expressed as kg/min. This unit was converted from a time-based measure of effort to a measure such as the number of fishing trips by dividing by the average length of a fishing trip.

Table 2: Estimates (tonnes) of illegal abalone catch, 2003–2007

Year	Catch
2003	5.8
2004	27.5
2005	68.7
2006	–
2007	89.0

Model equations

The exogenous parameters, values and references used are given in Table 1. The model equations are given in Table 3. The associated endogenous parameters are given in Table 4. The model comprises three components. The first component, the abalone population model, is a density-dependent logistic fisheries model. Fowler's density-dependent term was included to capture the effect of non-linear density dependence, following Milner-Gulland and Leader-Williams[6]. The logistic function has precedence in modelling South African abalone dynamics in Zone E[12], thereby enabling access to specific biological parameters for this growth function. However, no known study has modelled the non-linear density-dependence factor for South African abalone. But it is possible to use optimisation to estimate the value of the density-dependent term by minimising the difference between the model data and the historical data using the historical data published in Edwards and Plagányi[12], supplemented by the data in Table 1. The same technique is also used to estimate the value of the adjustment parameter n'.

The second component to the model is the fishing catch. Three types of fishing effort were present in the model over the historical period of the study (1980–2007). A recreational fishery was present during most of the time period, but was closed in 2004. A commercial fishery is also present in the study area, with the catch dependent on a TAC allocation determined annually by the fisheries authorities. In both these fisheries, catch is a function of the abalone abundance only. Naturally, however, recreational catches were set at zero from 2004 onwards. The illegal poaching, on the other hand, follows a standard Schaefer production function (Table 3). In other words, it is a function of the catchability coefficient, poaching effort and the available abalone population.

Table 3: Equations for the South African abalone fishery model

Cost for a diving expedition (US dollars per expedition):	**Revenue from sales of abalone confiscations:**
$c = p_C + p_B$	$\pi_t^M = b^D h_t^I V^A$
Payment to crew (US dollars per expedition):	**Illegal harvests in time t:**
$p_C = A_P^C Q_t^A$	$h_t^I = C_t^I X_t$
Quantity caught (kilograms per expedition):	**The decrease in poaching effort is:**
$Q_t^A = qvT_E X_t$ where v is a conversion rate from kilograms to tonnes	$\frac{dE}{dt} = n'E\left(\frac{c}{p} + \frac{bf}{p}\right)$
Payment to boat owners (US dollars per expedition):	**Growth in the abalone population is:**
$p_B = \frac{pc_B}{v}$	$\frac{dX}{dt} = srX$
Illegal catch coefficient:	**Removals from the abalone population:**
$C_t^I = mqT_E E_t$	$\frac{dX}{dt} = sX\left(\frac{rX^z}{k^z} + C_t^I + C^C + C_t^R\right)$
Commercial catch coefficient:	**Net present value of sales from confiscations:**
$C^C = \frac{\frac{1}{T_C}\sum_{t=1980}^{2007} h_t^C}{X_{1980}}$	$\pi(\delta,t) = \sum_{t=1}^{120} \frac{\pi_t^M}{(1+\delta)^n}$
Recreational catch coefficient:	**The increase in poaching effort:**
$C_t^R = \frac{\frac{1}{T_R}\sum_{t=1980}^{2003} h_t^R}{X_{1980}}$ $= 0.03,\ t \le 2003$ $= 0$ otherwise	$\frac{dE}{dt} = n'C_t^I X_t$ for $t \neq 2004$ $\frac{dE}{dt} = n'C_t^I X_t + C_R^I$ for $t = 2004$

Table 4: Endogenous variables in the model for the South African abalone fishery

Parameter / endogenous variable	Symbol	Unit
Commercial harvests in time t	h_t^c	Tonnes
Recreational catch coefficient in time t	C_t^R	Dimensionless
Illegal harvests in time t	h_t^I	Tonnes
Recreational harvests in time t	h_t^R	Tonnes
Poaching effort in time t	E_t	Number of expeditions
Abalone biomass in time t	X_t	Tonnes
Illegal catch coefficient in time t	C_t^I	Dimensionless
Payment given to boat owners	p_B	US dollars/expedition
Payment given to crew	p_C	US dollars/expedition
Quantity of abalone caught per expedition in time t	Q_t^A	Kilograms/expedition
Revenue from abalone confiscation in time t	π_t^M	US dollars/year
Enforcement costs for marine authority	c^M	US dollars/year

Illegal fishing is subject to enforcement effort by authorities. A third component to the model, therefore, is the authorities' response to illegal poaching. In standard deterrence models, there is one probability of detection that *prevents* illegal activity. However, in the South African abalone fishery, there are two detection probabilities, both of which are important to the model. The first probability is the one reported above, and the second is the probability of detection following an illegal poaching event. Once poaching occurs, abalone cannot be returned to the wild and an increase in the probability of detection after poaching therefore does not improve abalone populations. The seizure of illegally caught abalone by fisheries authorities is the most commonly reported on in the literature. In our model, therefore, there is both a probability of detection that prevents illegal harvesting, and a probability of detection that results in the seizure of abalone assets.

According to Goga[11], fishery authorities sell abalone seized after illegal poaching. There is therefore an incentive to rather improve the probability of detection *following* an illegal poaching event, than to improve the probability of detection *prior* to a poaching event. Another factor to consider in the model is the penalty for transgression. Although abalone is confiscated, an additional penalty is levied. Some studies report high penalties and prison sentences for offenders. However, a recent study in Hangberg[17] found that a fine levied to a repeat offender to prevent the possibility of a prison sentence was a mere ZAR20 000 (equivalent to USD2000 at the time at which the fine was imposed).

The marine authority's problem, according to neoclassical economic principles, is to maximise the present value of revenues from the confiscation of illegal catch:

$$PV = \int_0^{\infty} e^{-\delta t} b^D h_t^I(x,E) V^A dt, \qquad \text{Equation 5}$$

where δ is the discount rate, b^D is the probability of detection after a poaching event, $h_t^I(\bullet)$ is abalone confiscations and V^A is the market value of abalone sales.

The change in abalone population ($\dot{x}$) is a function of the population growth rate f(x) less harvests:

$$\dot{x} = f(x) - h^I(E,x) - h^L(x), \qquad \text{Equation 6}$$

where $h^I(E,x)$ is the catch from illegal harvesting which is a function of poaching effort (E) and stock size x, and $h^L(x)$ is the catch from legal harvesting which is only a function of stock size.

The change in poaching effort ($\dot{E}$) is a function of profitability from illegal harvesting $\pi(\bullet)$:

$$\dot{E} = \emptyset\pi(b,f;E,\bar{c},\bar{p}). \qquad \text{Equation 7}$$

The objective of marine authorities is to vary the probability of detection prior to a poaching event (b) and the magnitude of the fine (f) in order to maximise the present value of revenues (costs (c) and prices (p) are held constant). The system dynamics model is used to solve this problem, using a discrete time variant of the net present value problem (Table 3).

Results

Baseline simulation

The model captures the increase in the illegal poaching of abalone from 2003 onwards. This increase occurred despite greater enforcement of marine fisheries zones. In 2005, Operation Neptune and Operation Trident ended, as did the Environmental Courts. Although these initiatives were in the Overberg, poaching in the Table Mountain National Park doubled from 27.5 tonnes to 68.7 tonnes (Table 2). This model indicates that, for the given parameter values, poaching has continued to increase.

The model provides a reasonably good fit to the illegal catch, effort and abalone abundance time series between 1980 and 2007 (Figure 3a–e). Figure 3 indicates that illegal fishing spiked after the closure of the recreational fishery. The increase in illegal poaching (Figure 3b) mirrors the declines in recreational fishing (Figure 3a). Recreational fishing peaked at approximately 120 tonnes, whereas illegal fishing peaked at approximately 90 tonnes.

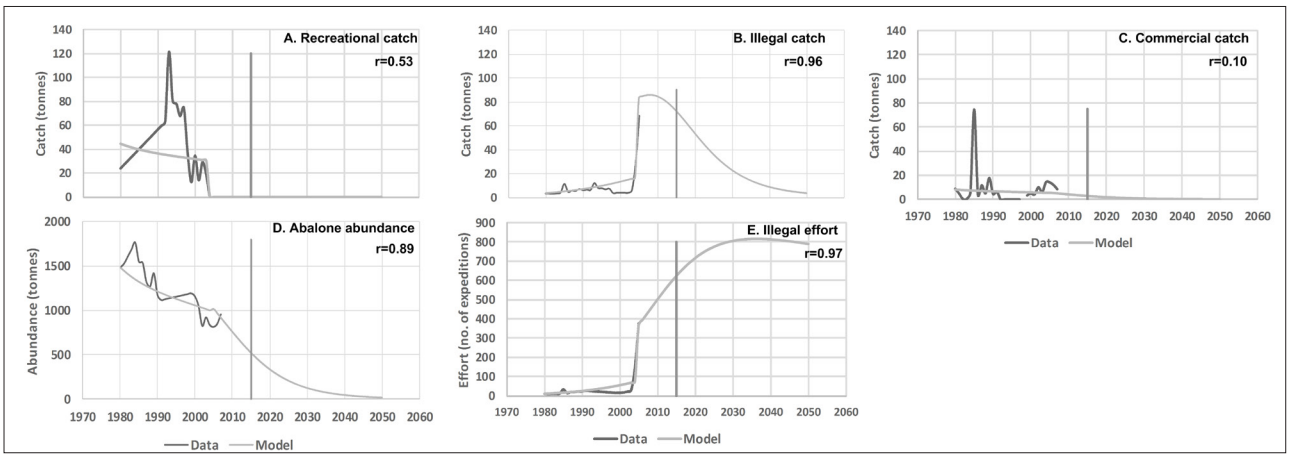


Figure 3: Baseline simulations showing how the model outputs of catch, abundance and effort replicate the historical data. The vertical line in each graph indicates the 2015 estimates.

Figure 3c indicates that, historically, the commercial fishery sectors have played a relatively small role in overall fishery catches. However, the most significant outcome of the model is that abalone stocks appear to have continued to decline since 2007, and are currently less than half of carrying capacity (Figure 3d). The decline in the illegal catch that is expected from 2015 onwards, it seems, is largely because of the lack of availability of abalone (Figure 3d) and spikes in poaching effort since 2003 (Figure 3e). According to the model, if current trends persist, abalone populations in Table Mountain National Park will be functionally extinct by 2050.

Policy simulations

The resource manager's challenge is to maximise the present value of abalone seizures by varying the probability of detection prior to a poaching event, as well as the fine for a transgression. One way of doing this is by patrolling no-take areas and also monitoring the departure of boats from the harbours.

It is important to emphasise that the following results hold for a resource manager's discount rate of less than 0.04 (4%). This is a model outcome. The marine authorities must place a high time preference on future stocks of abalone, in other words the sustainability of the resource. However, one would expect this from authorities that are custodians of abalone stocks.

The effect of varying the probability of detection prior to a poaching event (b) and the level of the fine (f) under costless enforcement are shown in Figure 4 in three figures: a bubble plot (Figure 4a), a line plot (Figure 4b) and an XY plot (Figure 4c).

The bubble plot (Figure 4a) indicates that the resource manager's profitability is maximised for high values of b and low values of f, and also low values of b and high values of f. Counter-intuitively, high values of both f and b result in the lowest net present values (NPVs) for resource managers, although these would result in the highest long-term abundance of abalone stocks.

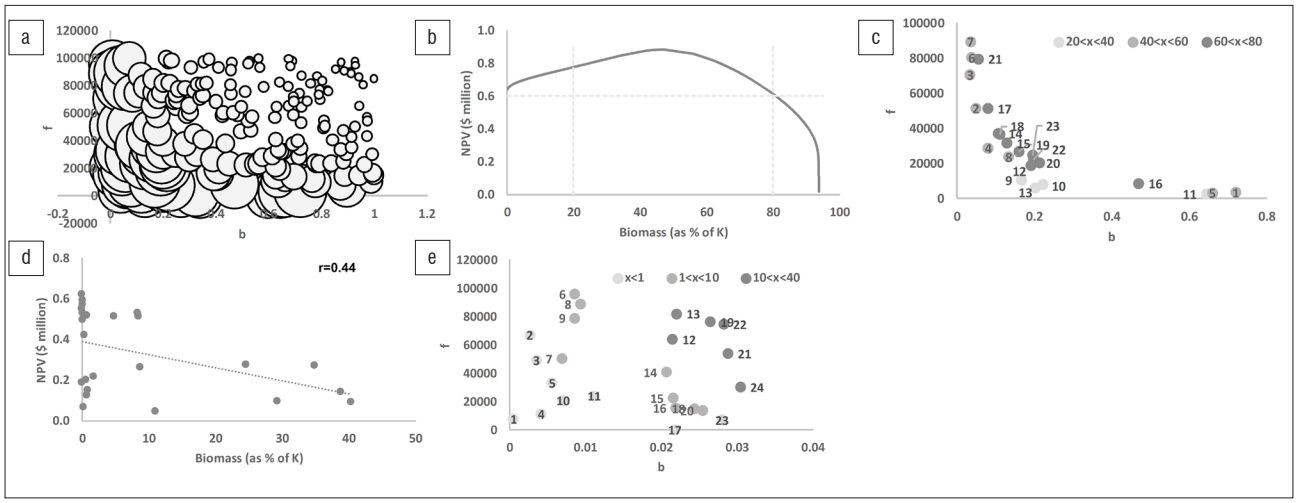
The line plot (Figure 4b) indicates that the marine authorities' NPV is maximised at an abalone abundance of approximately 50% of carrying capacity (0.47K). This value equates more or less to the stock density at maximum sustainable yield for the density-dependent logistic growth function (0.49K). The XY data plot of b and f (Figure 4c) indicates that the highest NPV is achieved for a high probability of detection (b) and a low fine (f). Taken together, this implies that improving the probability of detection prior to a poaching event produces a sustainable stock density in the abalone fishery, while also maximising profits for the marine authority.

The first series of policy simulations assumes that enforcement is costless. A simple cost function was added to the marine authority's decision problem:

$$c^M = bWc, \qquad \text{Equation 8}$$

where W is the enforcement effort of the marine authority. Enforcement costs are therefore an increasing function of the probability of detection prior to a poaching event (b). In the absence of other data, it is assumed that the enforcement costs per unit effort for the marine authority are the same as the cost per unit effort for the poaching vessel (c), and that the marine authority patrols once a day. Profit maximisation under this function is achieved by minimising the probability of detection prior to a poaching event, while maximising the level of the fine. The maximum payoff is USD0.63 million, compared with USD0.88 million under the zero enforcement cost outcome. However, at this level of enforcement effort the abalone population declines to zero by 2050, compared with populations stabilising at 47% of carrying capacity under the zero enforcement cost profit maximisation scenario.

As for the baseline, a Monte Carlo simulation was employed by varying the probability of detection prior to a poaching event (b) and the magnitude of the fine (f). This simulation indicated that only 3% of outcomes resulted in a positive NPV for the marine authority. Plotting these NPVs against biomass (from an ensemble of 750 realisations – Figure 4d) shows that, in contrast with the model of costless enforcement, profits for the resource authority are highest when stocks are allowed to deplete, and decline linearly as the biological population increases in relation to its carrying capacity. The XY plot under costly enforcement (Figure 4e) indicates that NPVs are highest under low probabilities of detection (b) and lowest under high probabilities of detection (although in absolute terms these probabilities of detection are much lower than under the costless enforcement scenario). Under costly enforcement, therefore, profit maximisation results in an unsustainable abalone population compared with profit maximisation under costless enforcement, and an incentive to minimise the probability of detection prior to a poaching event, compared with costless enforcement where there is an incentive to increase the probability of detection prior to a poaching event.



Note: (4a) Bubble size indicates NPV at year 2100; (4c and e) colour shades indicate abalone abundance as a percentage of carrying capacity (K), numbers indicate NPV rank, with 1 being the highest NPV.

Management variables: b is the probability of detection prior to an offence and f is the magnitude of the fine.

Figure 4: Simulations of the effect of management variables on marine authorities net present value (NPV) under (a–c) *costless* enforcement and (d,e) *costly* enforcement.

Sensitivity analysis

In order to test the sensitivity of the model outcomes to the new data on illegal catch in Zone E, the model was re-run using the illegal poaching data from Edwards and Plagányi[12] and the shift from recreational to poaching in the model was set at zero. The model outputs were then compared with the new historical data for abalone abundance and illegal poaching effort (Figure 5). Figure 5 indicates that the sensitivity analysis replicates the decline in abalone abundance relatively satisfactorily (Figure 5a), and tracks changes in effort relatively accurately until around 1992 (Figure 5b). Thereafter, the model predicts a continued increase in poaching effort, whereas the data show a decline in poaching effort from 1993 onwards.

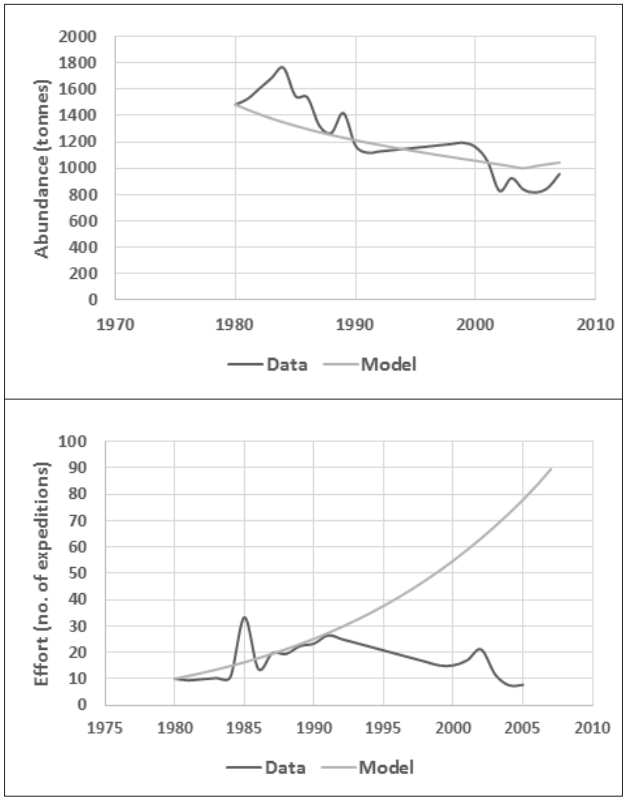


Figure 5: Sensitivity plots for (a) abalone abundance and (b) poaching effort.

There are therefore some discrepancies between the model and the published historical data under the sensitivity run. However, more recent data in Brill and Raemaekers[14] suggests that confiscations alone could be much higher than these published data. For example, Brill and Raemaeker's data indicate that reported confiscations in Zone E amounted to 17.2 tonnes of shucked abalone in 2005 (own calculations based on their data), which is considerably higher than the 1.4 tonnes reported in Edwards and Plagányi[12] for the same period. Furthermore Brill and Raemaeker's data exclude undetected poaching events, so the possibility exists that poaching is even higher in Zone E than these estimates. Their data span the years 2000 to 2009, and therefore do not include the period from 1993 to 1999. Their data do, however, indicate a growing trend in confiscations over these 9 years, although this would only indicate an increase in illegal catch (and poaching effort as effort is proportional to catch in the Schaefer production function – see Equation 2) over that period if b^D was stable or declining over that same period. Plagianyi et al.[18] published time series data of a policing effort index which could provide an indication of likely changes in b^D over time. The data show that the policing index remained constant between the years 2000 and 2004, but then decreased by 20 percentage points in 2004 and remained at this lower level until 2008. This finding indicates that illegal catch (and poaching effort) may have increased by even more than that modelled in this study (see Equation 4 – a decreasing b^D implies an increase in illegal catch and poaching effort). However, this index is for Zones A to D and one can therefore not categorically state that the same trends would be observed in Zone E. In conclusion, there still is much uncertainty over the historical data, and further work in this area would be beneficial.

In spite of these uncertainties, the sensitivity analysis indicates that abalone abundance will decline, reaching functional extinction by 2080, which is slightly longer than that for the baseline model (2050). The model is therefore robust in predicting declines in abalone abundance, irrespective of whether or not there is a spike in illegal catches following the closure of the recreational fishery.

Implications for rhino conservation

There are many similarities between the abalone fishery and rhino conservation in South Africa. Both abalone and rhino horn are high-value commodities for which there is considerable demand on the Asian markets. Although abalone can be sold legally for about ZAR350/kg (USD35/kg wet 'in shell' weight), black market prices are considerably higher, with recent estimates at approximately ZAR2500/kg (USD250/kg). However, the trade in abalone is high volume. It was recently estimated that the illegal trade in abalone in the Eastern Cape Province of South Africa alone was worth USD50–100 million annually.[19] The rhino horn trade fetches considerably higher prices, approximately USD65 000/kg in 2013, with some studies indicating prices as high as USD100 000/kg, but volumes are lower. Given that one horn weighs about 3 kg, and about 1000 rhinos were poached in South Africa in 2013, the implied illegal trade in rhino horn in South Africa was worth USD200–300 million in 2013.

Other similarities between the abalone fishery and rhino horn trade include that both products are relatively easy to transport; both species are slow-growing species and do not propagate easily; and as resources, they are difficult to protect – rhinos inhabit large areas that are often not adequately fenced off and abalone live in the ocean and consequently there are large areas of coastline to patrol. In sum, these types of resource demonstrate many features of an open-access system.

One notable difference between abalone and rhinos is that the abalone fishery in South Africa is characterised by a legal and illegal trade, whereas rhino horn is currently only traded illegally. However, there are strong lobby groups advocating for the legalisation of rhino horn trade. Observing the behaviour of the South African government as far as the South African abalone fishery is concerned therefore gives one some indication of how the government might respond if rhino horn were legally traded. There are, then, some important lessons that could be learnt from the abalone fishery.

The abalone fishery study indicates that it is important to distinguish between the probability of detection *prior* to the poaching offence, and the probability of detection *following* the poaching offence. Legalisation of trade in rhino horn would provide incentives to the South African government to increase the probability of detection *following* a poaching event. In the short term, this approach would not help to protect rhino populations, as the rhino would have already been killed. This approach will result in a race for rhinos and a 'trading on extinction' scenario in which profits are maximised over the short term to the detriment of the rhino population. This is similar to the 'banking on extinction' hypothesis of Mason et al.[20], except that instead of stockpiling, goods are sold on the open market. Over the long term, the effect of changing the probability of detection after a poaching event for rhinos is uncertain. However, changing the probability of detection after a poaching event for abalone fisheries has not deterred poachers over the long term. In fact, poaching appears to have increased over the periods of heightened enforcement effort between 2000 and 2005. The solution for escaping the 'trading on extinction' scenario is to improve the probability of detection *prior* to a poaching event.

Under costless enforcement, profit maximisation provides an incentive to reduce the probability of detection prior to a poaching event, which results in a sustainable resource stock approximately at maximum sustainable yield. However, under costly enforcement, there is no economic incentive for resource custodians to improve the probability

of detection prior to a poaching event. The abalone model indicates that the long-term sustainability of the resource is highly questionable under the following conditions: (1) a wildlife resource is subject to high levels of poaching; (2) there is a legal as well as an illegal trade in the resource; (3) the trade is highly lucrative; (4) profit maximisation principles are pursued by resource owners/custodians; and (5) enforcement is costly. Most of these conditions also hold for rhinos, were a legal trade to be allowed. A limitation of the abalone model is that only confiscation revenues, and not other sources of income, are included. For game farms or reserves, alternative sources of income such as that from de-horning and tourism may influence the outcome of the model. Some work has been done in this regard (see discussion section), although further work is required.

It is important also to recognise that a fundamental difference between abalone and rhinos is that abalone's primary benefit is through consumption. On the other hand, rhinos have a significant direct non-consumption benefit, namely tourism value. However both rhinos and abalone have a potential indirect non-consumptive benefit through ecosystem functioning. As Waldram et al.[21] indicate, rhinos may reduce sward height in grasslands, thereby reducing the extent of bushfires in savanna regions. Abalone, on the other hand, relies on the sea urchin *Parechinus angulosus* as a refugia for juveniles,[22] and therefore a well-functioning ecosystem is crucial to the survival of abalone. However, the removal of abalone may affect the emergence of epibenthic communities, as well as predatory fish diets.[23] Therefore, tourism in both the marine environment as well as the savanna areas may be affected indirectly through an overall degeneration of ecosystem functioning associated with the disturbance of the ecological system. An assessment of a legal trade in both rhino horn and abalone should therefore also take these non-consumptive benefits into consideration.

Discussion

It has been shown here that a trading on extinction incentive exists in the management of abalone stocks. This finding implies that there is an incentive for wildlife authorities to maximise short-term revenues at the expense of the long-term sustainability of the abalone population. A similar effect was observed by Crookes and Blignaut[24] in their model of the trade in rhino horn. They found that, although a legalisation of the trade increased the profitability of game farms significantly compared with a no-trade scenario, profits were higher under a policy outcome that resulted in the local extinction of the rhino population compared with profits under a sustainable supply of rhino. This outcome was a result of the fact that demand for rhino horn needed to decrease for sustainability to be achieved, which meant that the price of rhino horn would fall compared with demand from unsustainable supply.

It may seem paradoxical that a wildlife authority would have as its objective profit maximising, rather than another objective such as welfare maximisation, so two branches of literature are briefly discussed in support of this conclusion. Firstly, there is a significant body of literature that suggests that governments are motivated by profits when it comes to wildlife management. For example, Mayoral-Phillips, in his review of protected areas management in southern Africa, indicated that 'Governments, institutions, experts and stakeholders are rigidly following the rhetoric of "wildlife pays, wildlife stays", a form of wildlife privatisation of the commons'[25(p.11)]. This privatisation does not imply that a national resource is placed in private sector hands, but rather implies that private sector principles of profit maximisation are applied (for example see De Fraja and Delbono[26]).

Other studies have also confirmed that there is a detrimental effect on the environment if the principles of profit maximisation are applied by government in certain instances. Bárcena-Ruiz and Garzón[27] showed that when there are both government and private firms in product markets the environmental damage is greater than in the case of only private sector firms competing. Beladi and Chao[28] also found a detrimental effect on the environment if government follows profit maximising principles. This outcome is a consequence of the failure of the government sector to prevent environmental damage when it acts as a monopoly, because they argue that it is characterised by inefficient management (see also Wang and Wang[29]). In all these examples, however, environmental impact is measured in terms of pollution emissions. This study is the first one known to observe this trend in abalone, and there is now a growing body of literature to suggest that profit maximisation under a legalised trade in wildlife may be detrimental to the sustainability of the wildlife populations. This is true at least for rhinos and abalone, and further work is required to establish whether or not this conclusion holds in other wildlife sectors as well.

Conclusions

An open-access deterrence model for the abalone fishery in Table Mountain National Park (Zone E) was presented here. Fishing has moved westwards in recent years, because many areas of the Overberg region are now so depleted that some areas no longer contain viable populations. This study is therefore important, as it provides a tool for marine authorities to better manage abalone populations. Under current conditions, abalone populations in Table Mountain National Park could be depleted by 2050, as a result of a spike in illegal fishing effort following the closure of the recreational fishery in 2003.

In contrast with Gordon's seminal article[30], profit maximisation may not always result in the improved sustainability of wildlife populations. In the case of rhinos and abalone, a legalised trade must be accompanied by secondary policies, for the sustainability of the resource to be achieved. For rhinos, this secondary policy is consumer behaviour modification that reduces demand,[24] whereas for abalone, it is improving the probability of detection prior to an offence. Under costly enforcement, the market does not provide the incentives to adopt these secondary policies, because profits are higher without them. If there is a legalised trade then there is an incentive for resource owners to switch to a non-sustainable (but profit maximising) solution – termed the 'trading on extinction' effect. There is a need to emphasise and promote the non-consumptive benefits of the resources in question, as well as, in the case of abalone, address the causes of poaching rather than the consequences of poaching.

Acknowledgements

This article forms part of a 3-year project studying new approaches to model the economics of scarce resources. Funding from the National Research Foundation (South Africa) is gratefully acknowledged.

References

1. Muchapondwa E, Brick K, Visser M. Abalone conservation in the presence of drug use and corruption: Implications for its management in South Africa. Int J Sust Econ. 2014;6(2):201–216. http://dx.doi.org/10.1504/ijse.2014.060348
2. Raemaekers S, Hauck M, Bürgener M, Mackenzie A, Maharaj G, Plagányi EE, et al. Review of the causes of the rise of the illegal South African abalone fishery and consequent closure of the rights-based fishery. Ocean Coast Manage. 2011;54:433–445. http://dx.doi.org/10.1016/j.ocecoaman.2011.02.001
3. Hall RJ, Milner-Gulland EJ, Courchamp F. Endangering the endangered: The effects of perceived rarity on species exploitation. Conserv Lett. 2008;1:75–81. http://dx.doi.org/10.1111/j.1755-263X.2008.00013.x
4. Wilen JE. Common property resources and the dynamics of overexploitation: The case of the North Pacific fur seal. University of British Columbia Resources Paper No. 3. 1976; September.
5. Bjørndal T, Conrad JM. The dynamics of an open-access fishery. Can J Economics. 1987;20:74–85. http://dx.doi.org/10.2307/135232
6. Milner-Gulland EJ, Leader-Williams N. A model of incentives for the illegal exploitation of black rhinos and elephants: Poaching pays in Luangwa Valley, Zambia. J Appl Ecol. 1992;29:388–401. http://dx.doi.org/10.2307/2404508
7. Becker G. Crime and punishment: An economic approach. J Polit Econ. 1968;76(2):169–217. http://dx.doi.org/10.1086/259394
8. Van Heerden JH, Blignaut JN, Groenendijk NS. On the shady side of economics. S Afr J Econ Manag Sci. 1999;2(2):207–221.
9. Sutinen JG, Andersen P. The economics of fisheries law enforcement. Land Econ. 1985;61(4):387–397. http://dx.doi.org/10.2307/3146156
10. Kuperan K, Sutinen J. Blue water crime: Deterrence, legitimacy, and compliance, in fisheries. Law Soc Rev. 1998;32(2):309–338. http://dx.doi.org/10.2307/827765

11. Goga K. The illegal abalone trade in the Western Cape. ISS Paper 261. Pretoria: Institute for Security Studies; 2014.

12. Edwards CTT, Plagányi ÉE. Participatory assessment of the South African abalone resource and its impact on predicted population trajectories. S Afr J Sci. 2008;104:185–191.

13. Allen R, Punsly R. Catch rates as indices of abundance of yellowfin tuna, *Thunnus albacares*, in the eastern Pacific Ocean. Bull Inter-Amer Trop Tuna Comm. 1984;18(4):301–379.

14. Brill GC, Raemaekers SJPN. A decade of illegal fishing in Table Mountain National Park (2000–2009): Trends in the illicit harvest of abalone *Haliotis midae* and West Coast rock lobster *Jasus lalandii*. Afr J Mar Sci. 2013;35(4):491–500. http://dx.doi.org/10.2989/1814232X.2013.850443

15. Brandão A, Butterworth DS. Results for the abalone spatial- and age-structured assessment model for Zones A, B, C and D in 2013. Working Group document FISHERIES/2013/AUG/SWG-AB/12. Cape Town: Department of Agriculture, Forestry and Fisheries; 2013.

16. Hauck M, Gallardo-Fernández GL. Crises in the South African abalone and Chilean loco fisheries: Shared challenges and prospects. Mar Stud. 2013;12:3. http://dx.doi.org/10.1186/2212-9790-12-3

17. De Greef K. The booming illegal abalone fishery in Hangberg: Tough lessons for small-scale fisheries governance in South Africa [MSc dissertation]. Cape Town: University of Cape Town; 2013.

18. Plagányi É, Butterworth D, Burgener M. Illegal and unreported fishing on abalone: Quantifying the extent using a fully integrated assessment model. Fish Res. 2011;107:221–232. http://dx.doi.org/10.1016/j.fishres.2010.11.005

19. John Webb. Perlemoen poachers [television broadcast]. Carte Blanche. Johannesburg: M-Net; 2014 October 02.

20. Mason CF, Bulte EH, Horan RD. Banking on extinction: Endangered species and speculation. Oxford Rev Econ Pol. 2012;28(1):180–192. http://dx.doi.org/10.1093/oxrep/grs006

21. Waldram MS, Bond WJ, Stock WD. Ecological engineering by a mega-grazer: White rhino impacts on a South African savanna. Ecosystems. 2008;11:101–112. http://dx.doi.org/10.1007/s10021-007-9109-9

22. Blamey LK, Plagányi ÉE, Branch GM. Was overfishing of predatory fish responsible for a lobster-induced regime shift in the Benguela? Ecol Mod. 2014;273:140–150. http://dx.doi.org/10.1016/j.ecolmodel.2013.11.004

23. Blamey LK, Branch GM, Reaugh-Flower KE. Temporal changes in kelp forest benthic communities following an invasion by the rock lobster *Jasus lalandii*. Afr J Mar Sci. 2010;32(3):481–490. http://dx.doi.org/10.2989/1814232X.2010.538138

24. Crookes DJ, Blignaut JN. Debunking the myth that a legal trade will solve the rhino horn crisis: A system dynamics model for market demand. J Nat Conserv. 2015;28:11–18. http://dx.doi.org/10.1016/j.jnc.2015.08.001

25. Mayoral-Phillips AJ. Transboundary areas in southern Africa: Meeting the needs of conservation or development? Paper presented at: The Ninth Conference of the International Association for the Study of Common Property; 2002 June 17–21; Victoria Falls, Zimbabwe. http://dx.doi.org/10.1007/s10108-006-9006-y

26. De Fraja G, Delbono F. Alternative strategies of a public enterprise in oligopoly. Oxford Econ Pap. 1989;41(2):302–311.

27. Bárcena-Ruiz JC, Garzón MB. Mixed oligopoly and environmental policy. Span Econ Rev. 2006;8:139–160. http://dx.doi.org/10.1007/s10108-006-9006-y

28. Beladi H, Chao C-C. Does privatization improve the environment? Econ Lett. 2006;93:343–347. http://dx.doi.org/10.1016/j.econlet.2006.06.005

29. Wang LFS, Wang J. Environmental taxes in a differentiated mixed duopoly. Econ Syst. 2009;33:389–396. http://dx.doi.org/10.1016/j.ecosys.2009.08.002

30. Gordon HS. Economic theory of a common-property resource: The fishery. J Polit Econ. 1954;62:124–142. http://dx.doi.org/10.1086/257497

31. De Greef K, Raemaekers S. South Africa's illicit abalone trade: An updated overview and knowledge gap analysis. Cambridge: TRAFFIC International; 2014.

Angolan vegetable crops have unique genotypes of potential value for future breeding programmes

AUTHORS:
José P. Domingos[1]
Ana M. Fita[2]
María B. Picó[2]
Alicia Sifres[2]
Isabel H. Daniel[3]
Joana Salvador[3]
Jose Pedro[3]
Florentino Sapalo[3]
Pedro Mozambique[3]
María J. Díez[2]

AFFILIATIONS:
[1]Department of Biology, Science Faculty, University Agostinho Neto, Luanda, Angola
[2]Institute for the Conservation and Improvement of the Valentian Agrodiversity, Polytechnic University of Valencia, Valencia, Spain
[3]Center for Genetic Resources, University Agostinho Neto, Luanda, Angola

CORRESPONDENCE TO:
María José Díez

EMAIL:
mdiezni@btc.upv.es

POSTAL ADDRESS:
Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Edificio 8E, acceso J, 3er piso Camino de Vera s/n Valencia, Valencia 46022, Spain

DATES:
Received: 03 Aug. 2015
Accepted: 31 Oct. 2015

KEYWORDS:
vegetable genetic resources; molecular markers; African eggplants; cucurbits; phenotypic variability

HOW TO CITE:
Domingos JP, Fita AM, Picó MB, Sifres A, Daniel IH, Salvador J, et al. Angolan vegetable crops have unique genotypes of potential value for future breeding programmes. S Afr J Sci. 2016;112(3/4), Art. #2015-0285, 12 pages. http://dx.doi.org/10.17159/sajs.2016/20150285

A survey was carried out in Angola with the aim of collecting vegetable crops. Collecting expeditions were conducted in Kwanza-Sul, Benguela, Huíla and Namibe Provinces and a total of 80 accessions belonging to 22 species was collected from farmers and local markets. Species belonging to the Solanaceae (37 accessions) and Cucurbitaceae (36 accessions) families were the most frequently found with pepper and eggplant being the predominant solanaceous crops collected. Peppers were sold in local markets as a mixture of different types, even different species: *Capsicum chinense, C. baccatum, C. frutescens* and *C. pubescens.* Most of the eggplant accessions collected belonged to *Solanum aethiopicum* L. Gilo Group, the so-called 'scarlet eggplant'. Cucurbita genus was better represented than the other cucurbit crops. A high morphological variation was present in the *Cucurbita maxima* and *C. moschata* accessions. A set of 22 *Cucurbita* accessions from Angola, along with 32 *Cucurbita* controls from a wide range of origins, was cultivated in Valencia, Spain and characterised based on morphology and molecularity using a set of 15 microsatellite markers. A strong dependence on latitude was found in most of the accessions and as a result, many accessions did not set fruit. The molecular analysis showed high molecular variability and uniqueness in the collected accessions, as shown by their segregation from the set of global controls. In summary, the material collected is quite valuable because of its uniqueness and the potential of the breeding characteristics it possesses.

Introduction

Angola is a country with enormous agricultural potential, with vast cultivable and underutilised areas, excellent water resources and a large variety of climates and ecosystems. However, huge challenges still exist with regard to food security and the high levels of poverty among the rural population. Approximately half the Angolan population live in rural areas, and more than 85% of growers are small farmers that cultivate crops for self-consumption. Landraces adapted to their specific agro-climatic conditions and requiring low inputs are maintained by these growers. These varieties are quite valuable as they present unique adaptations to specific cultivation systems and soil types.

The typical Angolan diet is based on the consumption of various legumes, cereals and roots, such as cassava, millet, sorghum, beans, sweet potato, peanut, rice, wheat and potatoes. The cultivation of vegetable crops is scarce, which contributes to a low intake of vitamins and minerals among the population. Some initiatives are currently being developed to increase the cultivation of vegetables. One example is the urban and peri-urban agriculture implemented in the neighbourhood of Luanda. This is playing a crucial role in ensuring the population's access to high-quality fresh food. However, in order to increase the benefits of this and other initiatives, and facilitate a more market-oriented production system, it is crucial that they employ more suitable varieties that will maximise yields and adapt to specific conditions. At present, Angola has alternative varieties of vegetable crops such as pumpkin, eggplant, pepper and watermelon. However, little information is available regarding the genetic resources of these crops or of the adaptability of breeding cultivars to suit specific Angolan growing conditions.

In 1991, The Center for Plant Genetic Resources of Agostinho Neto University (CRF-UAN) in Luanda started a programme with the purpose of collecting and conserving the genetic resources of food-plant landrace species. The CRF-UAN currently holds a collection of 4100 accessions, of which only 8.8% are horticultural vegetable crops. All of the accessions are preserved under long-term conditions with the aim of characterising them and incorporating them into breeding programmes. Owing to economic restrictions and a lack of technicians, only 7% of all the accessions have been characterised further, and only 4% have been assessed for genetic diversity using molecular markers. Of these characterised accessions, none have been horticultural crops.

The CRF-UAN is aware of the importance of conducting detailed studies of the collection, as their utilisation in plant breeding requires a profound knowledge of their morphological, agronomical and genetic characteristics. A joint project between the Instituto de Conservación y Mejora de la Agrodiversidad Valenciana of the Universitat Politècnica of València (COMAV-UPV) and Agostinho Neto University in Luanda was initiated in 2012. The project was funded by the Agencia Española de Cooperación Internacional. The global aim of this project was to strengthen the use of Angolan vegetable genetic resources, with the understanding that in order to do so, it would be necessary to further characterise these resources. A second objective was to provide a fully equipped molecular laboratory to the Centre for Genetic Resources of Luanda and to train its staff in collecting, morphological characterisation and molecular techniques, as well as in the use of vegetable genetic resources in breeding programmes.

We describe a collecting expedition conducted in 2012 in four Angolan provinces with the aim of enriching the vegetable collection of the CRF-UAN of Luanda. Additionally, we conducted a molecular study of the collected *Cucurbita* material, using microsatellite markers, to obtain information about the genetic variability of the Angolan *Cucurbita* accessions and their relatedness to a set of reference accessions from all over the world.

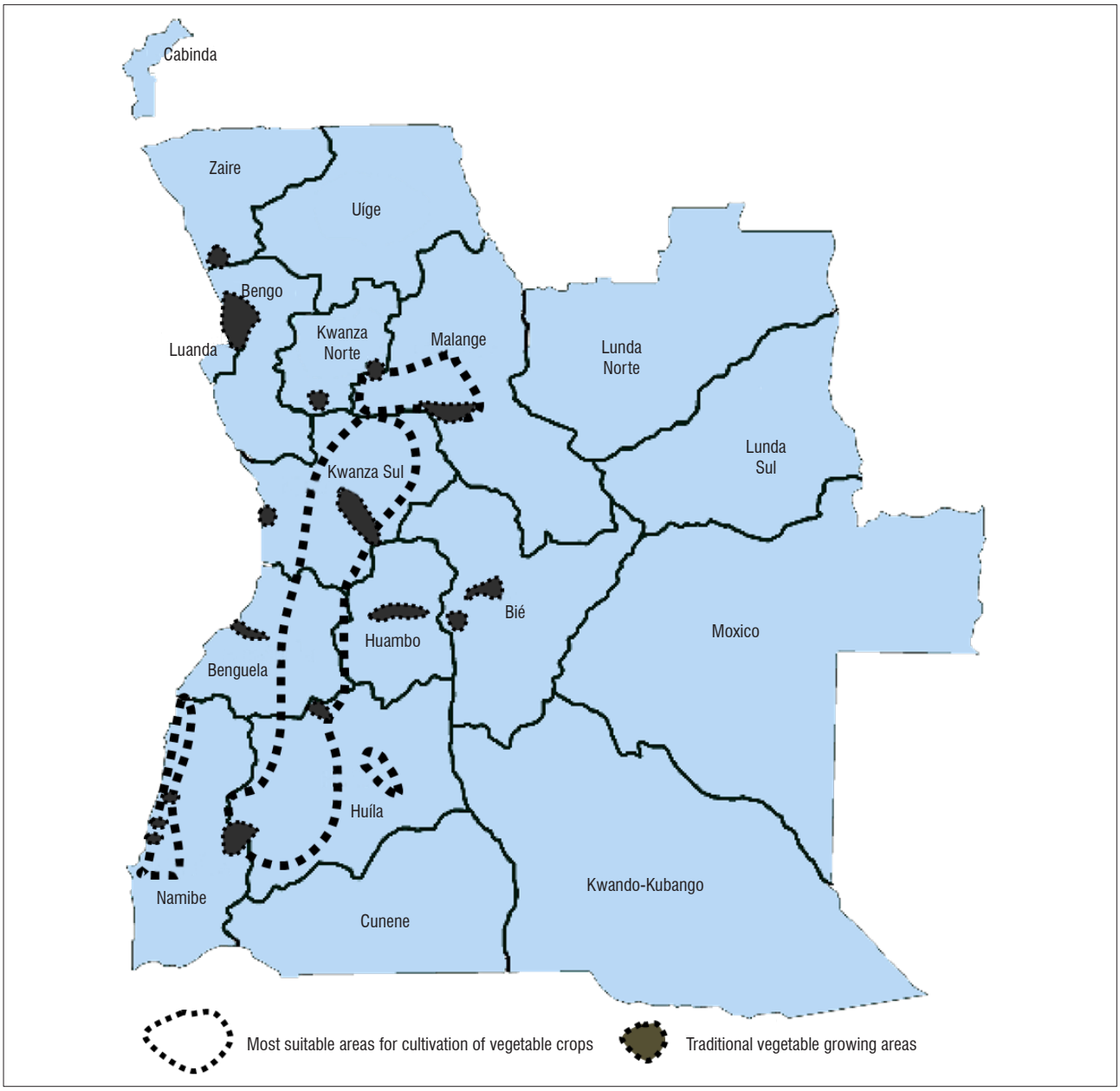


Figure 1: Areas visited during the collecting expedition.

Material and methods

Areas surveyed

The provinces of Kwanza-Sul, Benguela, Huíla and Namibe were all visited during the collecting expedition. These areas are part of the seven traditional areas of vegetable production, and are among the 5 most suitable areas for the cultivation of these types of crops in Angola. All 4 provinces are situated along the western strip at altitudes between 1000 m and 1200 m, and the period of exploitation of crops lasts about 6 months, from May to October[1] (Figure 1). According to Diniz[1], Lombe, in Malanje Province, and Cela, in Kwanza-Sul Province, both at an altitude of 1200 m, supply vegetable products during the dry seasons and even for part of the rainy season. Other growing areas are distributed throughout the centre and south, especially in the areas connected by the Benguela railway (Catabola, Chinguar, Bela Vista and Vila Verde) and in the southern regions of Lubango-Humpata-Chibia and Matala-Capelongo.

Kwanza-Sul encompasses the vegetable-producing areas of the south shore of the Kwanza river (Libolo-Amboim) and the transitional central-northern region that comprises a large part of the strip that has a tropical, hot and semi-arid climate, with a rainy season of about seven months (October-April). The provinces of Benguela, Huíla and Namibe belong to four agricultural areas: the south shore, the transitional central-western region, the transitional central-southern region and Gambos, where the climate is similar to the humid heat of the interior, semi-arid high plateau, according to Thornthwaite's classification.[2]

The collecting expedition was conducted jointly by Angolan and Spanish researchers. In all provinces, previous contacts established with the local agricultural agencies facilitated direct contact with the farmers and the collection of accessions on their own farms. Other accessions were collected in local markets. In all cases, passport data included in the Multicrop Passport Descriptor List[3] were recorded for each collected accession. The Centre for Plant Genetic Resources of Luanda holds seed samples of all the collected accessions. Some were sent to the Genebank of the COMAV-UPV for further characterisation and molecular analysis in collaboration with the Angolan researchers.

Table 1: *Cucurbita* spp. accessions collected in Angola and used in the morphological and molecular diversity assay

Species	Accession (code used in Figures 3–6)	Collection site	Latitude	Longitude	Altitude (m)	Collection source
C. maxima	ANG01 (ma01)	Porto Amboin, Kwanza-Sul	101135S	0133117E	34	Local market
C. maxima	ANG08 (ma08)	Ebo, Kwanza-Sul	105123S	0143847E	1192	Local market
C. maxima	ANG11 (ma11)	Kibala, Kwanza-Sul	103842S	0145747E	1387	Local market
C. maxima	ANG12 (ma12)	Kibala, Kwanza-Sul	103842S	0145747E	1387	Local market
C. maxima	ANG13 (ma13)	Kibala, Kwanza-Sul	103842S	0145747E	1387	Local market
C. maxima	ANG14 (ma14)	Kibala, Kwanza-Sul	103842S	0145747E	1387	Local market
C. maxima	ANG17 (ma15)	Kibala, Kwanza-Sul	103842S	0145747E	1387	Local market
C. maxima	ANG18 (ma18)	Kibala, Kwanza-Sul	103842S	0145747E	1387	Local market
C. maxima	ANG38[a] (ma38)	Baia Farta, Benguela	125651S	0130534E	7	Seeds from farmer
C. maxima	ANG44 (ma44)	Baia Farta, Benguela	123832S	0131638E	7	Local market
C. maxima	ANG50 (ma50)	Cacula, Huíla	141507S	0140223E	1577	Orchard
C. maxima	ANG52 (ma52)	Cacula, Huíla	143258S	0140439E	1577	Bought from road vendor
C. maxima	ANG53 (ma53)	Cacula, Huíla	143258S	0140439E	1577	Bought from road vendor
C. maxima	ANG54 (ma54)	Cacula, Huíla	143258S	0140439E	1577	Bought from road vendor
C. maxima	ANG73 (ma73)	Chibia, Huíla	151242S	0134423E	1527	Local market
C. maxima	35040 (ma35040)	Kwanza-Sul	–	–	–	Local market
C. moschata	ANG03 (mo03)	Gabela, Kwanza-Sul	105211S	0142538E	1077	Small village
C. moschata	ANG04 (mo04)	Ambuíla, Kwanza-Sul	105211S	0142538E	1077	Small village
C. moschata	ANG15[a] (mo15)	Kibala, Kwanza-Sul	103842S	0145747E	1387	Local market
C. moschata	ANG16 (mo16)	Kibala, Kwanza-Sul	103842S	0145747E	1387	Local market
C. moschata	35039 (mo35039)	Kwanza-Sul	–	–	–	Local market
Cucurbita pepo	ANG58[a] (pe58)	Humpata, Huíla	150237S	0133025E	1872	Small village

[a] *Accessions tested with the complete set of microsatellites*

Plant material

Accessions of 22 different species were collected. Those belonging to the genus *Cucurbita* were used for the genetic diversity analysis. In total, 22 *Cucurbita* accessions from Angola were assayed, along with 32 *Cucurbita* spp controls from a wide range of origins, representing most of the cultivated and wild species of the genus.[4] Seeds of the control accessions were supplied by the United States Department of Agriculture (USDA), the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), the COMAV and the Vavilov Institute of Plant Industry (VIR) genebanks (Tables 1 and 2).

Morphological characterisation

A total of 22 accessions from Angola (three plants per accession) were cultivated in greenhouses at the UPV. According to a first tentative classification based on seed morphology and fruit pictures taken during the collection, all of these accessions except one belonged to the species *Cucurbita maxima* and *C. moschata* (Table 1). The following plant and flower descriptors were recorded: days to male/female flowering (measured as the number of days from sowing to the opening of the first male/female flowers), node in which the first male/female flower appeared, days to fruit setting (number of days from sowing to the first fruit set) and presence of leaf spots (0 absence, 1 presence).

Isolation of DNA and Simple Sequence Repeat analysis

DNA was extracted from the young leaves of 3 plants per accession using the Cetyl Trimethyl Ammonium Bromide method with minor modifications (Table 3).[5] DNA concentrations were measured with a NanoDrop 1000 Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA) and adjusted to 50 ng/μl. Twenty-six microsatellite *loci* were selected from those previously identified in an *in silico* screen of an Expressed Sequence Tag (EST) collection of *C.pepo*.[6] These Simple Sequence Repeats (SSRs) were previously validated in a *C.pepo* collection,[7] and in the present study were further validated in a wide range of *Cucurbita* species. First, the amplification of the whole set of SSRs was checked in a subset of 5 accessions belonging to various accessions of the main cultivated species (*C.pepo*, *C.moschata* and *C.maxima*), including the parentals of the *C. pepo* genetic map from which these SSRs were derived.[5] All the SSRs that amplified in the three species and which indicated variability were analysed in the full collection. Amplifications were carried out in 15 μl reactions containing 25 ng template DNA, polymerase buffer, 2 mM $MgCl_2$, 200 mM dNTPs, 160 nM labelled forward primer, 200 nM unlabelled reverse primer and 0.5 U Taq polymerase. Polymerase chain reaction (PCR) was performed using a cycling profile of one cycle at 94 °C for 3 min, one cycle at 94 °C for 30 s, 65 °C for 30 s and 72 °C for 30 s, plus 9 cycles of amplification with a descending annealing temperature of 1 °C per cycle, followed by 20 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s with a final extension at 72 °C for 10 min. The forward primers

were 5' labelled with FAM, PET, HEX and NED to allow for amplicon detection after separation on an ABI 3100 capillary DNA sequencer (Applied Biosystem Inc, Foster City, CA, USA).

Data analysis

A Factorial Correspondence Analysis (FCA) was conducted with the SSR data using the Genetix 4.05.2 programme.

Results

Accessions collected in the visited areas

A total of 80 accessions were collected: 13 in Benguela, 30 in Huíla, 32 in Kwanza-Sul, and 3 in Namibe (Table 4). It was not possible to explore the Bibala or Camacuio municipalities, the most important agricultural areas of the Namibe province. Even though no collecting activity was planned in Luanda, 2 accessions were collected in this province.

Table 2: *Cucurbita* spp accessions from different origins included as controls in the molecular analysis

Species	Accession (code used in Figures 3–6)	Country of origin	Source
C. argyrosperma	PI 438547 (argy47)	Belize	ARS-GRIN
C. argyrosperma	PI 451712 (argy12)	USA	ARS-GRIN
C. cordata	PI 653839 (cord)	Mexico	ARS-GRIN
C. ecuadorensis	PI 432443 (ecua)	Ecuador	ARS-GRIN
C. foetidisima	PI 442191 (foet)	Mexico	ARS-GRIN
C. lundelliana	PI 540898 (lund)	Honduras	ARS-GRIN
C. maxima	AN-CU-59 (maSP)	Spain	COMAV-UPV
C. maxima	CATIE-9824 (maCO)	Colombia	CATIE
C. maxima	SUD-CU-6 (maAR)	Argentina	COMAV-UPV
C. maxima	VAV-1860 (maAU)	Australia	VIR
C. moschata	AFR-CU-1 (moMO)	Morocco	COMAV-UPV
C. moschata	AN-CU-45 (moSP)	Spain	COMAV-UPV
C. moschata	CATIE-7223 (moPA)	Panama	CATIE
C. moschata	KUROKAWA (moJA)	Japan	COMAV-UPV
C. moschata	NIGERIAN LOCAL (moNI)	Nigeria	COMAV-UPV
C. moschata	PI 264551 (moGU)	Guatemala	ARS-GRIN
C. moschata	PI 369346 (moCR)	Costa Rica	ARS-GRIN
C. moschata	PI 381814 (moIN)	India	ARS-GRIN
C. moschata	PI 419083 (moCH)	China	ARS-GRIN
C. moschata	PI 458746 (moGU)	Guatemala	ARS-GRIN
C. moschata	PI 482527 (moZB)	Zimbabwe	ARS-GRIN
C. moschata	PI 498429 (moCO)	Colombia	ARS-GRIN
C. moschata	PI 512150 (moMX)	Mexico	ARS-GRIN
C. moschata	PI 543218 (moBO)	Bolivia	ARS-GRIN
C. okeechobensis ssp. martinezii	PI 512105 (okee)	Mexico	ARS-GRIN
C. pedatifolia	PI 442290 (peda)	Mexico	ARS-GRIN
C. pepo	CATIE-11869 (peGU)	Guatemala	CATIE
C. pepo	UPV-196[a] (peSPV)	Valencia, Spain	COMAV-UPV
C. pepo	MU-CU-16[a] (peSPM)	Murcia, Spain	COMAV-UPV
C. pepo var *fraterna*	PI 532354 (peMX)	Mexico	ARS-GRIN
C. pepo var *ozarkana*	PI 614701 (peUS)	USA	ARS-GRIN
C. scabridifolia	PI 532391 (scab)	Mexico	ARS-GRIN

[a] *Accessions tested with the complete set of microsatellites*

Table 3: Microsatellites tested in five accessions of *C.pepo*, *C.maxima* and *C.moschata* (marked with the letter 'a' in Table 1 and Table 2).

Unigene in transcriptome version 1[a]	Unigene/SSR number in transcriptome version 3[b]	Expected size (bp)	Observed size (bp)	Amplification	Polymorphic
CUTC004307	CUUC107629/ SSR2993	100	120-145	yes	Yes
CUTC001906	CUUC87397/ SSR 1124	198	200-255	only *C. pepo*	only *C. pepo*
†CUTC004158	CUUC96784/ SSR 1755	188	200-204	yes	yes
CUTC005739	CUUC85492/ SSR 1020	194	200-204	only *C. pepo*	only *C. pepo*
CUTC004399	CUUC62112/ SSR 147	165	175-200	yes	no
CUTC004782	–	164	175-200	yes	no
CUTC005800	CUUC64302/ SSR 397	189	200-240	no	unclear
†CUTC006703	CUUC61907/ SSR 122	123	145	yes	yes
†CUTC008357	CUUC109093/ SSR 3236	136	156	yes	yes
†CUTC006209	CUUC119994/ SSR 5054	200	200-300	yes	yes
CUTC018879	CUUC98150/ SSR 1863	164	175-200	yes	no
CUTC012342	CUUC97483/ SSR 1806	182	200-204	yes	?
†CUTC009316	CUUC104381/ SSR 2559	300	320	yes	yes
†CUTC009760	CUUC62185/ SSR 2870	159	179	yes	yes
CUTC022867	–	140	145-175	yes	unclear
CUTC011336	–	149	145-175	no	no
†CUTC017708	–	223	250-300	yes	yes
†CUTC002749	CUUC65488/ SSR 603	183	175-200	yes	Yes
†CUTC020992	CUUC96401/ SSR 1725	156	175	yes	yes
†CUTC023363	–	244	264	yes	yes
†CUTC008722	–			yes	yes
†CUTC007942	CUUC118813/ SSR 4820	152	175-200	yes	yes
†CUTC006891	CUUC96843/ SSR 1759	154	145-175	yes	yes
†CUTC004991	CUUC114609/ SSR 4102	173	175-200	yes	yes
CUTC009607	CUUC108020/ SSR 3041	146	145-200	yes	?
†CUTC008409	–	107	120-145	yes	yes

[a] These SSR loci were previously validated in a C.pepo collection in Formisano et al.[7] This column shows the name of each SSR according to this publication (based on a v1 assembly of the C.pepo transcriptome available in the cucurbigene.net database).

[b] This column indicates the unigene name in which each SSR marker is located and the SSR number according to a new assembly of the C.pepo transcriptome (v3 available at cucurbigene.net). Details about the gene annotation, repeated motif, etc. are available in the cucurbigene database.

†SSR selected for genotyping the whole collection

Some of the accessions were later subdivided as they were composed of a seed mixture of different types, primarily of those belonging to the *Capsicum* genus. Most of the accessions belonged to the Solanaceae (37 accessions) and Cucurbitaceae (36 accessions) families, as they were the focus of the project. The same solanaceous and cucurbit types were found repeatedly in different local markets and in farmers' fields. To avoid too much redundancy, we collected only representative accessions of each type.

During the trip, we gained a good idea of the current status of horticulture in these provinces. After many years of continuous conflict, many of the traditional varieties that were the focus of this project, mainly solanaceous and cucurbitaceous crops, had become extinct. This was especially evident in tomato, melon, watermelon and cucumber crops. Only landraces of *Cucurbita* species and some types of African eggplants continued to be cultivated extensively. *Cucurbita maxima* and *C. moschata* were the most abundant species, although we cannot disregard the cultivation to a higher extent of certain species of *Lagenaria* genus in other areas that we were unable to visit.

The most predominant types of pepper, locally called 'guindungo',[8] were the small-sized ones, some of which were only 1 cm long. They are sold in local markets as a mixture of different types, including various species, mainly *Capsicum chinense, C. baccatum, C. frutescens* and *C. pubescens* (Figure 2). This is why almost half of the pepper accessions were subdivided into different types. Most of the 12 eggplant accessions collected belonged to the *Solanum aethiopicum* L. Gilo Group, the so-called 'scarlet eggplant' and 2 accessions belonged to the wild species *S. lichtensteinii*.

Table 4: Areas surveyed and number of accessions collected

Province (total number of accessions collected)	Localities	Species (Number of accessions)
Luanda (2)	Luanda	*Citrullus lanatus* var *citroides* (2)
Kwanza-Sul (32)	Honga, Zambia, Kondé, Demba, Kibala, Kizenza, Gabela	*Arachis hipogea* (2), *Capsicum* spp. (8), *Cucurbita* spp. (15), *Lagenaria siceraria* (1), *S. aethiopicum* (4), *S. melongena* (1), *S. lycopersicum* (1)
Benguela (13)	Catumbela, Dombe Grande, Bahía Azul, Bahía Farta, Chongoroi	*Capsicum* spp. (5), *Cucumis sativus* (1), *Cucurbita* spp. (3), *Solanum melongena*, *S. aethiopicum* and spp. (3), *Solanum lycopersicum* var. *cerasiforme* (1)
Huíla (30)	Sonso, Cacula, Matala, Lubango, Huíla-Humpata, Mulenga, Ontimpe, Chibia, Strada de Leba, Sierra de Leba	*Allium cepa* (1), *Capsicum* spp. (5), *Citrullus* var citroides (3), *Cucumis anguria* (1), *Cucurbita* spp. (7), *Daucus carota* (1), *Lactuca sativa* (1), *Lagenaria siceraria* (1), *S. aethiopicum* (2), *S. lycopersicum* (1), *S. l.* var. *cerasiforme* (3), *Solanum lichtensteinii* (2), *Vigna unguiculata* (2)
Namibe (3)	Guirau de Lima, Kanbongue	*Capsicum* spp. (1), *Citrullus lanatus* var *lanatus* (1), *Cucumis melo* (1)

Figure 2: Accessions collected in Angola. From left to right and from to top to bottom: pepper: ANG74, ANG22, ANG33, ANG34, ANG45, ANG75; eggplant: ANG24, ANG31, ANG42; C moschata: ANG04, ANG09, ANG15; C. maxima ANG12, ANG17, ANG73, ANG14, ANG50; C. pepo: ANG58.

A few accessions of *S. lycopersicum* var *cerasiforme* were also collected. Some of these, which were round or plum-shaped, were bought in small local markets; others were found growing as weeds close to commercial tomato fields. Five accessions of *Citrullus lanatus* var *citroides* were also collected in Luanda and Huíla Provinces.

Morphological description of the Cucurbita *spp. accessions collected*

The collected accessions of *Cucurbita* showed high variability in the morphological characteristics of their fruits, such as fruit size, shape (flattened, round, oblong or elongated), rind colour (white, cream, green or orange) and presence/absence of ribs, stripes and spots (Figure 2). Flesh colour was also variable (yellow or from light- to dark-orange).

The cultivation of the plants at the UPV permitted the taxonomic classification of the accessions. Depending on the seed, leaf and flower morphologies, the presence/absence of spots on the leaves and the plant growth habit, it was possible to classify the accessions as *C. maxima, C. moschata* or *C. pepo.* The peduncle morphology was also studied in the originally collected fruit to confirm the specific classification of the *Cucurbita* accessions. Accession ANG37 was composed of a mixture of *C. moschata* and *Lagenaria siceraria* seeds and was separated into two accessions. Many of the accessions cultivated in the greenhouses of the UPV did not set fruits, as a result of a very delayed flowering, probably because of not being adapted to Spain's latitude (39°28´11˝N). This was more noticeable in the *C. moschata* species, in which 35% of the plants did not flower, compared to 0,05% in the case of *C. maxima* (Table 5). A total of 4 accessions of *C. maxima* developed female flowers, which allowed the regeneration of three accessions, whereas only one accession of *C. moschata* developed female flowers. Most of the other accessions of the COMAV, USDA, CATIE and VAVILOV collections, which came from different origins and belonged to these and other species (*C. pepo, C.maxima* and *C.moschata)*, produced both male and female flowers under the same conditions, showing their complete adaptation after several years of cultivation in Valencia. In general, the fruits obtained from the Angolan accessions did not look like the original fruits collected. Also, the different plants of the same accession that set fruit revealed the existence of some segregation for morphological traits of the fruits. This means that the collected accessions are heterogeneous and they belong to highly heterozygous populations, probably due to cross-pollination between plants cultivated by different growers.

Genetic diversity of Cucurbita *species*

In the FCA performed with the SSR results, the *Cucurbita* accessions from Angola were mostly clustered into two groups with the reference accessions of *C. maxima* and *C. moschata*, thus confirming the morphological results that assigned most of the accessions to these two species (Figure 3). The accession ANG58, belonging to the species *C. pepo*, is located between the *maxima* and *moschata* groups. The rest of the accessions appeared disperse and more distant.

Table 5: Number of days from transplant to male and female flowering in Angolan accessions (mean of three plants)

Species	Accession	Number of days to male flowering	Node first male flower	Number of days to female flowering	Node first female flower
C. maxima	ANG01	45.33	10.33	47	23.00
C. maxima	ANG08	59.50	21.00	91	22.00
C. maxima	ANG11	61.67	19.70	No flowering	–
C. maxima	ANG12	58.67	19.70	No flowering	–
C. maxima	ANG13	76.00	20.00	No flowering	–
C. maxima	ANG14	67.67	22.00	No flowering	–
C. maxima	ANG17	155.00	23.50	No flowering	–
C. maxima	ANG18	61.33	16.30	No flowering	–
C. maxima	ANG38	50.67	15.00	59	26.70
C. maxima	ANG44	57.67	23.70	No flowering	–
C. maxima	ANG50	55.50	23.50	No flowering	–
C. maxima	ANG52	54.33	20.00	No flowering	–
C. maxima	ANG53	74.67	19.00	No flowering	–
C. maxima	ANG54	72.50	20.50	No flowering	–
C. maxima	ANG73	55.67	18.70	77	24.00
C. maxima	35040	56.50	23.50	No flowering	–
C. moschata	ANG03	58.50	21.50	No flowering	–
C. moschata	ANG04	115.50	23.00	No flowering	–
C. moschata	ANG15	76.50	23.00	No flowering	–
C. moschata	ANG16	No flowering	–	No flowering	–
C. moschata	35039	87.67	22.00	107	31.00
C. pepo	ANG58	59[2]	20.00	64	38.00

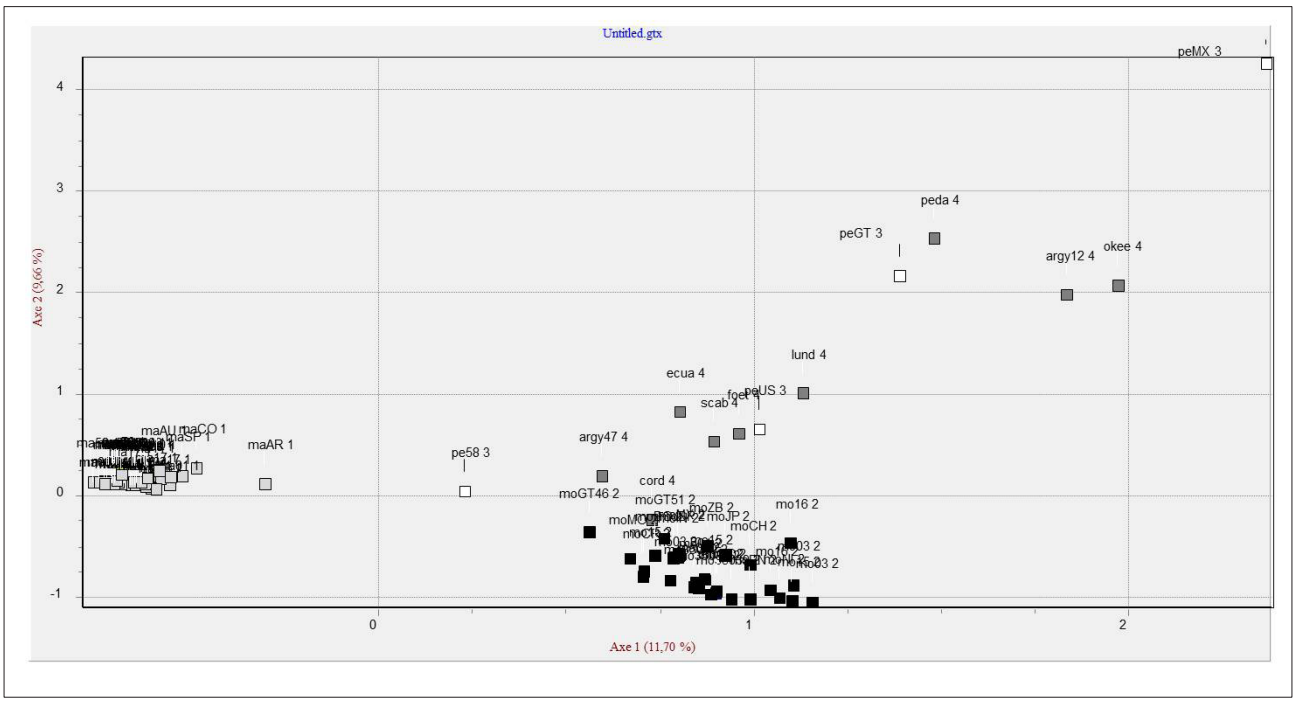


Figure 3: FCA of the Angolan and reference Cucurbita accessions representing the entire Cucurbita genus. Black squares: C. moschata, light grey: C. maxima, white squares: C. pepo and dark grey squares: other species. See tables 1 and 2 for accession codes.

Performing the analysis with *C. maxima* and *C. moschata* separately highlighted the segregation between the two as well as the greater variability of *C. moschata* compared to *C. maxima,* as detected with this set of SSRs (Figure 4). When analysed separately, the accessions of *C. moschata* were highly diverse, to the point of appearing scattered (Figure 5). Both the intra- and inter-accession variabilities were remarkably high in the Angolan germplasm, as can be seen on the right half of the figure. It is separate from the other accessions used as controls, although it does show continuous variation. The most distant accessions from the Angolan ones are those of eastern origin (China, India and Japan) as well as some from Central America (Mexico and Guatemala). In contrast, the closest accessions are from Africa, mainly Nigeria and Morocco, along with others from South America (Colombia and Bolivia) and the southern part of Central America (Costa Rica and Panama), which may suggest a Central to South American origin of these African accessions. It is interesting to note that the *C. moschata* accessions from Angola were just as diverse as were the reference accessions from many different regions of the world.

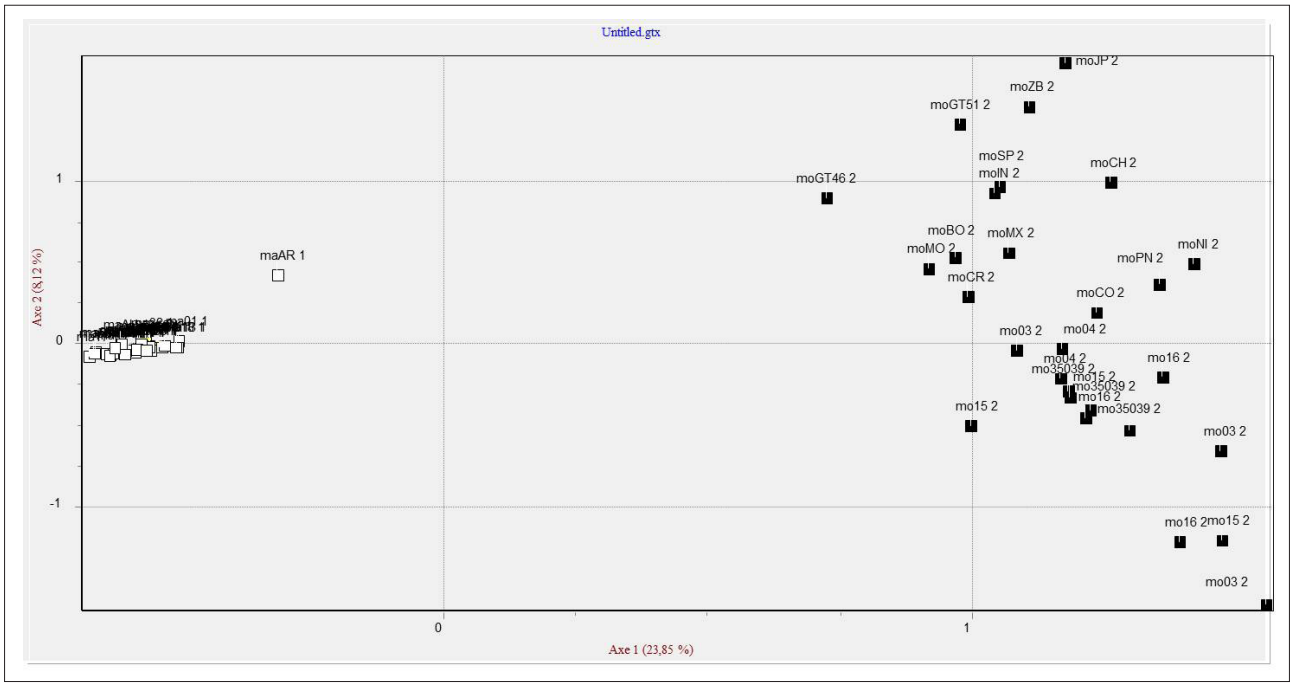


Figure 4: FCA of the C. maxima and C. moschata accessions. Black squares: C. moschata, white squares: C. maxima. See tables 1 and 2 for accession codes.

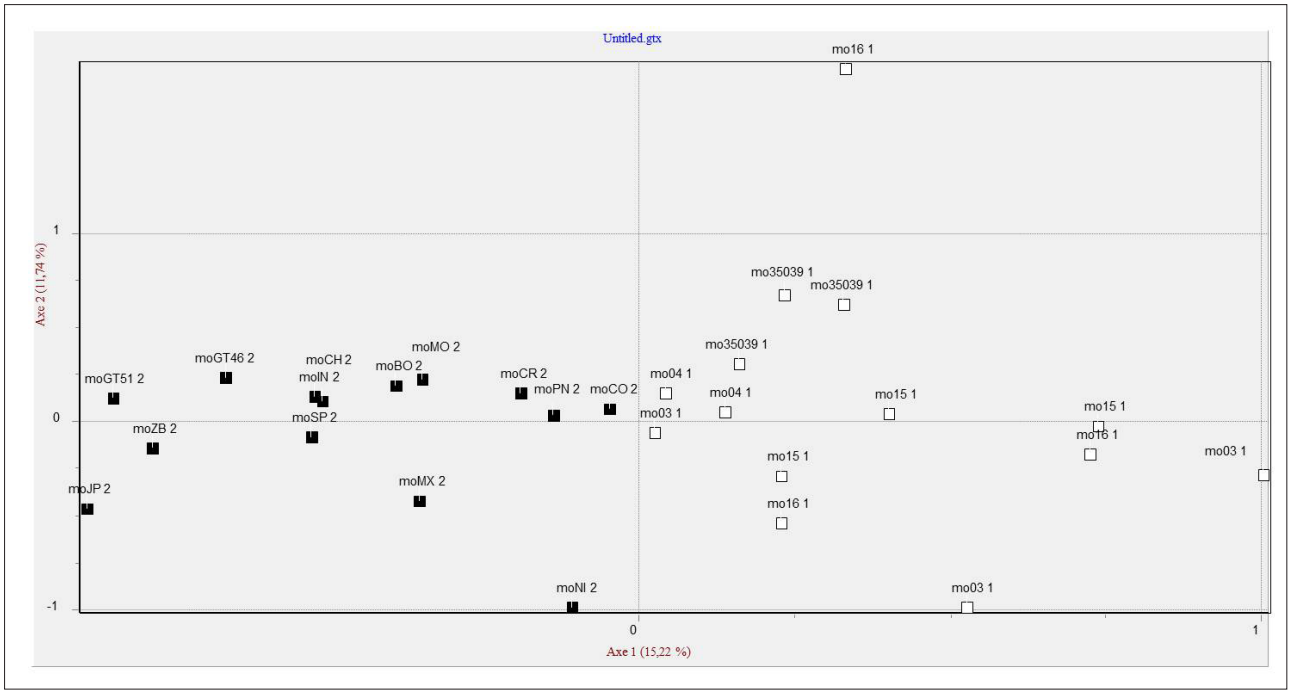


Figure 5: FCA performed with the Angolan and control accessions of C. moschata. Black squares: controls, white squares: Angolan accessions. See tables 1 and 2 for accession codes.

The accessions of *C. maxima* seem to be much less diverse, as most of them grouped together except for ANG17, ANG12 and ANG13 (Figure 6). On the other hand, the Angolan accessions were further away from those used as controls (VAV1860 and CATIE9824). The controls SUD-CU-6 and AN-CU-59 were excluded from the analysis because they were located quite distant from the others, which made the complete set of accessions studied too closely grouped. FCA seems to separate the accessions from coastal and northern Kwanza-Sul from those from interior and southern Huíla.

The described results of the FCA agree with the polymorphism analysis (Table 6). Therefore, when considering only *C. maxima* and *C. moschata* accessions, the highest polymorphism values, largest number of alleles per locus and highest Polymorphism Information Content (PIC) values of the microsatellites correspond to the Angolan accessions of *C. moschata*, followed by the Angolan *C. maxima,* whereas the *C. moschata* controls from other origins were less variable.

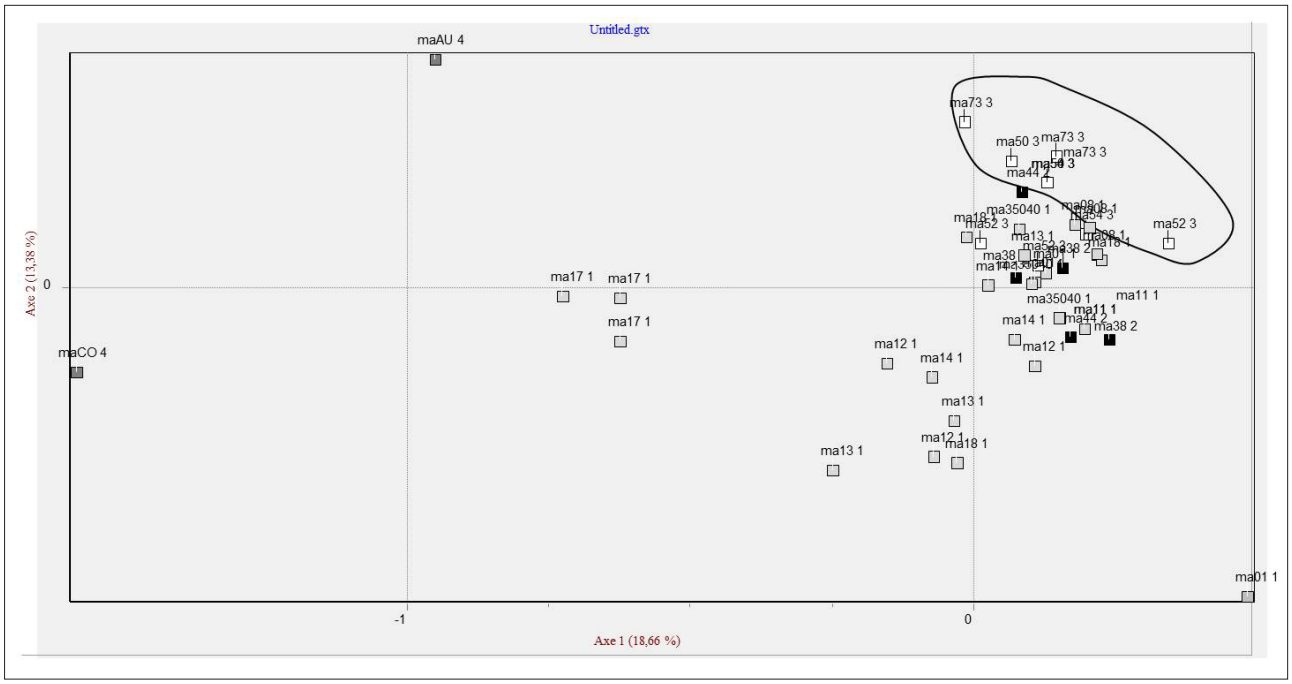


Figure 6: FCA performed with the Angolan and controls accessions of C. maxima. Black squares: accessions from Benguela, light grey squares: accessions from Kwanza Sul, white squares: accessions from Huila. Accession from Huila included in the circle. See tables 1 and 2 for accession codes.

Table 6: Estimates of molecular polymorphism of the accessions studied by species and origin

Species	Origin (number of plants)	Observed heterozygosity	Polymorphism at 0.95 and 0.99 levels	Number of alleles per *locus*	Polymorphic Information Content
C. maxima	Angolan (n=43)	0.0578	0.40/0.47	1.47	0.12
C. moschata	Angolan (n=14)	0.0758	0.47/0.60	1.60	0.17
	Controls (n=14)	0.0324	0.29/0.36	1.27	0.15
C. maxima and *C. moschata*	All accessions including controls of *C. maxima* (n=75)	0.0578	0.87/0.93	3.00	0.33
All accessions studied	Angolan and controls (n=88)	0.0634	0.93/1.00	4.8	0.43

Discussion

A collecting expedition was carried out in Angola in 2012 by the Centre for Plant Genetic Resources of Luanda and the Univesitat de València, Spain, with the goal of improving the collection of vegetable crops at the Centre for Plant Genetic Resources of Luanda. Morphological and molecular characterisation of part of the collected accessions was also conducted jointly by the Spanish and Angolan researchers. This expedition was part of a wider project aimed at developing new varieties of vegetable crops that are more productive and better adapted to the agro-climatic conditions of the visited Angolan provinces. The final, long-term goal was to promote the cultivation of the improved varieties of vegetables and introduce them into the Angolan diet, which will ultimately contribute to their diversification and increase their nutritional value.

A total of 80 accessions of vegetable crops were collected, most of which belonged to the *Cucurbita, Solanum* and *Capsicum* genera. We were able to confirm the existence of subsistence agriculture and a slow emergence of commercial agricultural production in the visited areas – to a higher extent in the northern provinces and to a lesser degree in Namibe, a more barren province which relies on fishing.

Squashes, pumpkins, peppers and eggplants were found to be the most commonly cultivated crops. Interestingly, pumpkins and eggplants are crops with low farm work needs. Likewise, small-sized peppers need fewer inputs and less care from growers, and set a higher number of fruits much more easily than plants with larger fruits. These types of crops seem to be the ones that have a higher probability of success in unsophisticated agricultural systems like the ones used in the visited areas. Other crops of the same families, such as tomato, melon or cucumber, with more specific cultivation and climatic needs, are much less likely to grow appropriately and give cost-efficient yields to farmers.

Most of the pepper accessions collected belonged to the *C. chinense, C. frutescens, C. baccatum* and *C. pubescens* species. They were small-sized and sold in local markets as a mix of different types. Given the elevated amount of trade between Angola and Brazil, we speculate that these pepper types may have come from Brazil. A molecular analysis of the collected accessions, including some from South America for reference, could help to elucidate the origin of these materials. In any case, these accessions are of great interest for two reasons: their potential to provide genes resistant to diseases and their adaptation to the specific agricultural conditions of these provinces. In fact, many genes of resistance introgressed in commercial cultivars come from this species.[9]

African eggplants are common in local markets (Figure 2). Most of the twelve eggplant accessions collected belonged to the *Solanum aethiopicum* L. Gilo Group, the so-called 'scarlet eggplant'. This species, together with *S. macrocarpon* (called gboma), are cultivated African eggplants that are locally important in their region of origin in tropical sub-Saharan Africa.[10-13] *S. aethiopicum* is a hypervariable species characterised by many morphologically different types and forms, with hundreds of local varieties.[14,15] Both scarlet and gboma eggplants are also important genetic resources for common eggplant (*S. melongena* L.) breeding, as the three species can be intercrossed, thus producing hybrids with intermediate fertility.[16] Two accessions of the very variable *Solanum lichtensteinii* were also collected in Huíla province. This species grows as a weed in disturbed and overgrazed areas, along roadsides and is used as a traditional medicine. The most outstanding value of this eggplant collection is its uniqueness and specific adaptation to the climatic conditions of Angola.

A total of 5 accessions of the wild form of watermelon *Citrullus lanatus* var. *citroides* were collected. This old variety is perfectly adapted to arid and desert conditions and plays a major role in the economy of the San of the Kalahari Desert. It accumulates a large amount of water in its fruits and is commonly used both to feed livestock as well as for human consumption during the dry season, especially the non-bitter form. In fact, it has been identified as a species with potential for commercial exploitation in arid areas of Botswana.[17] Additionally, *C. lanatus* var. *citroides* is readily crossable with the cultivated watermelon, which makes it of great interest for breeding, as it has been found to be resistant to different species of nematodes.[18,19]

The morphological characterisation conducted with the *Cucurbita* accessions was our first contact with the Angolan materials and allowed us to classify most of them as *C. maxima* and *C. moschata* species. Although the genus is native to the northern parts of South America,[20] after their domestication, most species spread to the rest of the world, and secondary centres of diversification exist with landraces adapted to different conditions.[2] The centre of diversity of *C.maxima* lies in the South American temperate zones, but many landraces of this species from other regions, including many countries in tropical Africa (the countries neighbouring Angola, such as the Democratic Republic of Congo, Zambia, Zimbabwe, Nigeria and Cameroon), are also maintained in international germplasm collections. In addition, the greatest diversity occurs among the Central and South America landraces of *C.moschata*,[21] but there also exists a significant diversity in landraces from the warm regions of other continents, including Africa and Asia.[22,23] In fact, tropical pumpkin (*C.moschata*) is one of the most important vegetable crops in some African countries.[24] In general, *C. maxima* is more important in the cooler parts of southern Africa and the Sahel region, and less important in west and east Africa, where it is more humid, and where *C. moschata* is more common. Our results fit with this information. The explored area of Angola is located between the latitudes 10° and 15° south, where the humid climate and tropical forest give way to different types of savannah, the predominant ecosystem of Angola, which is more welcoming to *C. maxima* species. The fruits, leaves and seeds of the *Cucurbita* species are consumed extensively in Angola. They are locally called 'abóbora', 'cabaça', 'diangua', 'dinhungo', 'matira' and 'malengue'. Croquettes of crushed seeds are used to cook a meat or fish stew called 'muteta'. Leaves (locally called miengueleka) are also stewed together with the fruit of the palm (*Elaeis guineense*) or groundnut (*Arachis hypogaea*) and salt.

The fruits are used to cook soup, puree and cakes. Other *Cucurbita* species also exist as landraces for farmers' own consumption and local markets. These crops contribute to nutrition through their abundant supply of minerals.

The agronomic characterisation and molecular analysis conducted have unveiled certain aspects of the cultivation requirements of these plants and also the genetic variability of the Angolan accessions. The collected materials are strongly adapted to the specific agro-climatic conditions of the collection areas. This has been revealed by the lack of flowering when cultivated over a long period (from February to December) at the latitude of Valencia, Spain. Many *Cucurbita* landraces from tropical areas are not adapted to temperate regions, especially the many cultigens of *C. moschata* that are short-day plants which require long growing seasons.[25,26] Fewer adaptation problems arose with *C. maxima.*[27,28] The behaviour observed in our assay agrees with this differential of specific requirements.

The molecular analysis conducted demonstrated high molecular variability and uniqueness in the collected accessions, as shown by the segregation of the collection from the analysed set of controls from all over the world. *C. moschata* is the most diverse species, even more so, in terms of number of alleles per locus and PIC value, than the set of controls included in the assay from a wide range of origins. Also, the collected accessions of *C. maxima* show high levels of variability, and seem to group according to their origin, suggesting an adaptation of the different types to different agro-climatic areas. There are only a few papers prior to this one that describe the molecular variation in the African germplasm of these species, but none describes the diversity in Angola. For example, Gwanama et al.[29] analysed a collection of *C.moschata* accessions from Zambia and Malawi using RAPDs, and also found differences related to the origin of the accessions.

Our molecular results showed a high intra-accession variability, suggesting that cross-pollination probably occurs in their natural habitat. Also, the proximity of different fields in their traditional mode of cultivation most likely favours the existence of cross-pollination between different types.

The variability and singularity of the accessions makes this material highly valuable for use in breeding and even more so seeing as previous analyses of African accessions of *C. moschata* indicate resistance to viruses and fungi.[30] For example, Nigerian Local, an accession from Nigeria, has been reported as resistant to potyvirus and powdery mildew and has already been used for *Cucurbita* breeding. The accessions from Angola seem to be genetically close to this material, so they must be further analysed to determine their potential for pest and disease resistance. Also, some of the collected accessions have yellow-to-dark-orange flesh. Intensity of flesh colour is known to be related to the carotenoid composition of the fruit, mainly α and β-carotene, β-criptoxanthina, lutein and zeaxanthin,[31] so this material could be a promising source of vitamins in this country. Screening of accessions has been conducted in order to identify accessions with high carotenoid and ascorbic acid contents.[32] Some screenings have used landraces from secondary centres of diversity, such as Brazil and Malaysia, and have found high levels of variation for these traits in landraces from different areas.[33-35] Further studies will be conducted with the set of collected accessions to evaluate their carotenoid content. Accessions with high carotenoid contents may be utilised for further quality improvement of pumpkin during the development of new varieties, which will contribute to minimising nutritional problems in the population, especially those arising from vitamin A deficiency in children and women.

In summary, the material collected in this expedition is quite valuable given its uniqueness and its desirable breeding characteristics. Accessions from Angola are scarce in germplasm collections all over the world. As an example, none of the 72 accessions of *Solanum aethiopicum* included in the European Cooperative Programme's eggplant database come from this country. The same is true for the 67 accession of this species included in the USDA germplasm databases. In fact, Angola has been avoided by extensive collecting expeditions for many years, probably because of security reasons. These collected accessions can be used in breeding for many purposes, including resistance to diseases, adaptation to abiotic stresses, such as arid conditions, and to improve fruit quality. In the case of the *Cucurbita* accessions, the high molecular variability found makes the use of this adapted material in breeding quite promising. Breeding programmes aimed at increasing yield, resistance to diseases and quality may be successful using this starting material. However, complementary to breeding programmes, other initiatives focused on improving cultivation techniques should be undertaken to take advantage of all the genetic potential of the improved varieties.

Acknowledgements

This work, project A1/039611/11, was funded by the Agencia Española de Cooperación Internacional para el Desarrollo (Spanish Agency of International Cooperation for the Development).

Authors' contributions

MJD, JPD, PM and MBP conceived and designed the study; AMF, JP, FS and MJD participated in the collecting expedition; AMF, AS, IHD, JS and NBP performed the molecular analysis, MBP designed and performed the phenotyping trial; and MJD and MBP wrote the manuscript. All authors critically reviewed and approved the final manuscript.

References

1. Diniz AC. Angola o Meio Físico e Potencialidades Agrárias [Angola, physical environment and agricultural potential]. Lisboa: Ministério dos Negócios Estrangeiros; 1998. Portuguese.
2. Diniz AC. Características Mesológicas de Angola: descrição e correlação dos aspectos fisiográficos, dos solos e da vegetação das zonas agrícolas angolanas [Mesological characteristics of Angola: description and correlation of physiographic aspects, soils and vegetation of agricultural Angolan areas]. 2nd ed. Lisboa: Missão de Inquéritos Agrícolas de Angola; 2006. Portuguese.
3. Alercia A, Diulgheroff S, Mackay M. Food and Agriculture Organization of the United Nations (FAO)/Bioversity multi-crop passport descriptors V.2. Rome: Bioversity International; 2012.
4. Esteras C, Nuez F, Picó B. Genetic diversity studies in Cucurbits using molecular tools. In: Behera T, Wang Y, editors. Cucurbits. Series: Genetics, genomics and breeding in crop plants. Enfield, NH: Science Publishers; 2011. p. 140–198. http://dx.doi.org/10.1201/b11436-6
5. Esteras C, Gómez P, Monforte AJ, Blanca J, Vicente-Dólera N, Roig C, et al. High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait loci mapping. BMC Genomics. 2012;22(13):80. http://dx.doi.org/10.1186/1471-2164-13-80
6. Blanca JM, Cañizares J, Roig C, Mir G, Ziarsolo P, Nuez F, et al. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). BMC Genomics. 2011;12:104. http://dx.doi.org/10.1186/1471-2164-12-104
7. Formisano G, Roig C, Esteras C, Ercolano MR, Nuez F, Monforte AJ, et al. Genetic diversity of Spanish *Cucurbita pepo* landraces: An unexploited resource for summer squash breeding. Genet Resour Crop Ev. 2012; 59:1169–1184. http://dx.doi.org/10.1007/s10722-011-9753-y
8. Ribas O. Alimentação Regional Angolana [Regional Angolan diet]. Luanda: Edição Centro de Informação e Turismo de Angola; 1992. Portuguese.
9. Crosby KM. Pepper. In: Prohens J, Nuez F, editors. Handbook of plant breeding. Vegetables II. New York: Springer; 2008. p. 221–248. http://dx.doi.org/10.1007/978-0-387-74110-9_6
10. Lester RN, Jaeger PML, Bleijendaal-Spierings BHM, Bleijendaal HPO, Holloway HLO. African eggplants – a review of collecting in West Africa. FAO/IBPGR Plant Genet Resour Newsl. 1990;81–82:17–26.
11. Schippers RR. African indigenous vegetables. An overview of the cultivated species. London: University of Greenwich, Natural Resources Institute; 2000.
12. Lester RN, Daunay MC. Diversity of African vegetable *Solanum* species and its implications for a better understanding of plant domestication. Schriften zu Genetischen Ressour. 2003; 22:137–152.

13. Maundu P, Achigan-Dako E, Morimoto Y. Biodiversity of African vegetables. In: Shackleton CM, Pasquini MW, Drescher AW, editors. African indigenous vegetables in urban agriculture. London: Earthscan; 2009. p. 65–104.

14. Lester RN, Hakiza JJH, Stavropoulos N, Teixeira MM. Variation patterns in the African scarlet eggplant, *Solanum aethiopicum* L. In: Styles BT, editor. Infraspecific classification of wild and cultivated plants. Oxford: Clarendon Press; 1986. p. 283–307.

15. Plazas M, Andújar I, Vilanova S, Gramazio P, Herraiz FJ, Prohens J. Conventional and phenomics characterization provides insight into the diversity and relationships of hypervariable scarlet (*Solanum aethiopicum* L.) and gboma (*S. macrocarpon* L.) eggplant complexes. Front Plant Sci. 2014;5:318. http://dx.doi.org/10.3389/fpls.2014.00318

16. Prohens J, Plazas M, Raigón MD, Seguí-Simarro JM, Stommel JR, Vilanova S. Characterization of interspecific hybrids and backcross generations from crosses between two cultivated eggplants (*Solanum melongena* and *S. aethiopicum* Kumba group) and implications for eggplant breeding. Euphytica. 2012;186:517–538. http://dx.doi.org/10.1007/s10681-012-0652-x

17. Taylor FW. The potential for the commercial utilization of indigenous plants in Botswana. In: Wickens GE, Goodin JR, Field DV, editors. Plants for arid lands. London: George Allen and Unwin Ltd; 1985. p. 231–242. http://dx.doi.org/10.1007/978-94-011-6830-4_17

18. Thies JA, Levi A. Resistance of watermelon germplasm to the peanut root-knot nematode. HortScience. 2003;38:1417–1421.

19. Thies JA, Levi A. Characterization of watermelon (*Citrullus lanatus* var. *citroides*) germplasm for resistance to root-knot nematodes. HortScience. 2007;42:1530–1533.

20. Sanjur OI, Piperno DR, Andres TC, Wessel-Beaver L. Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. P Natl Acad Sci USA. 2002;99:535–540. http://dx.doi.org/10.1073/pnas.012577299

21. Wessel-Beaver L. Evidence for the Center of Diversity of *Cucurbita moschata* in Colombia. Cucurbit Gen Coop Report. 2000;23:54–55.

22. Andres TC. Diversity in tropical pumpkin (*Cucurbita moschata*): A review of infraspecific classifications. In: Lebeda A, HS Paris, editors. Progress in cucurbit genetics and breeding research. Proceedings of Cucurbitaceae. 8th EUCARPIA meeting on cucurbit genetics and breeding; 2004 July 12-17; Olomouc, Czech Republic: Olomouc: Palacký University; 2004. p. 107–112.

23. Wu J, Chang Z, Wu Q, Zhan H, Xie S. Molecular diversity of Chinese *Cucurbita moschata* germplasm collections detected by AFLP markers. Sci Hortic-Amsterdam. 2012:128(1):7–13. http://dx.doi.org/10.1016/j.scienta.2010.12.006

24. Gwanama C, Botha AM, Labuschagne MT. Genetic effects and heterosis of flowering and fruit characteristics of tropical pumpkin. Plant Breeding. 2001;120:271–272. http://dx.doi.org/10.1046/j.1439-0523.2001.00595.x

25. Ferriol M, Picó B, Fernández de Córdoba P, Nuez F. Molecular diversity of a germplasm collection of squash (*Cucurbita moschata*) with SRAP and AFLP markers. Crop Sci. 2004;44(2):653–664. http://dx.doi.org/10.2135/cropsci2004.6530

26. Ferriol M, Picó B. Pumpkin and winter squash. In: Prohens J, Nuez F, editors. Handbook of plant breeding. Vegetables I. New York: Springer; 2007. p. 317–349.

27. Ferriol M, Picó B, Nuez F. Genetic diversity of some accessions of *Cucurbita maxima* from Spain using RAPD and SRAP markers. Genet Resour Crop Ev. 2003;50(3):227–238. http://dx.doi.org/10.1023/A:1023502925766

28. Ferriol M, Picó B, Nuez F. Morphological and molecular diversity of a collection of *Cucurbita maxima* landraces. J Am Soc Hortic Sci. 2004;129(1):60–69.

29. Gwanama C, Labuschagne MT, Botha AM. Analysis of genetic variation in *Cucurbita moschata* using random amplified polymorphic DNA (RAPD) markers. Euphytica. 2000;113:19–24. http://dx.doi.org/10.1023/A:1003936019095

30. Paris HS, Brown RN. The genes of pumpkin and squash. HortScience. 2005;40:1620–1630.

31. Tadmor Y, Paris HS, Meir A, Schaffer AA, Lewinshon E. Dual role of the pigmentation gene B in affecting carotenoid and vitamin E content in squash (*Cucurbita pepo*) mesocarp. J Agric Food Chem. 2005;53:9759–9763. http://dx.doi.org/10.1021/jf0520591

32. Pandey S, Singh J, Upadhyay AK, Ram D, Rai M. Ascorbate and carotenoid content in an Indian collection of pumpkin (*Cucurbita moschata* Duch. ex Poir.). Cucurbit Gen Coop Report. 2003;26:51–53.

33. Jaeger de Carvalho LM, Barros P, De Oliveira RL, Pacheco S, Fernandes do Monte PH, Viana JL, et al. Total carotenoid content, α-carotene and β-carotene, of landrace pumpkins (*Cucurbita moschata* Duch): A preliminary study. Food Res Int. 2012;47:337–340. http://dx.doi.org/10.1016/j.foodres.2011.07.040

34. Norshazila S, Irwandi J, Othman R, Yumi Zuhanis HH. Scheme of obtaining ß-carotene standard from pumpkin (*Cucurbita moschata*) flesh. Int Food Res J. 2012;19:531–535.

35. Norshazila S, Irwandi J, Othman R, Yumi Zuhanis HH. Carotenoid content in different locality of pumpkin (*Cucurbita moschata*) in Malaysia. Int J Pharm Pharm Sci. 2014;6:29–32.

Changing boundaries: Overcoming modifiable areal unit problems related to unemployment data in South Africa

AUTHOR:
Gina Weir-Smith[1,2]

AFFILIATIONS:
[1]Research Data Management Centre, Human Sciences Research Council, Pretoria, South Africa

[2]Honorary Research Fellow, School of Geography, Archaeology & Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Gina Weir-Smith

EMAIL:
gweir-smith@hsrc.ac.za

POSTAL ADDRESS:
Research Data Management Centre, Human Sciences Research Council, Private Bag X41, Pretoria 0001, South Africa

DATES:
Received: 13 Mar. 2015
Revised: 12 June 2015
Accepted: 16 Nov. 2015

KEYWORDS:
data aggregation; unemployment; areal interpolation; socio-economic trends; census

HOW TO CITE:
Weir-Smith G. Changing boundaries: Overcoming modifiable areal unit problems related to unemployment data in South Africa. S Afr J Sci. 2016;112(3/4), Art. #2015-0115, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20150115

The longitudinal comparison of census data in spatial format is often problematic because of changes in administrative boundaries. Such shifting boundaries are referred to as the modifiable areal unit problem (MAUP). This article utilises unemployment data between 1991 and 2007 in South Africa to illustrate the challenge and proposes ways to overcome it. Various censuses in South Africa use different reporting geographies. Unemployment data for magisterial districts of census 1991 and 1996 were re-modelled to the 2005 municipal boundaries. This article showed that areal interpolation to a common administrative boundary could overcome these reporting obstacles. The results confirmed more accurate interpolations in rural areas with standard errors below 3300. Conversely, the largest errors were recorded in the metropolitan areas. Huge increases in unemployment between 1996 and 2001 statistics were also evident, especially in the metropolitan areas. Although such areas are more complex in nature, making it more difficult to accurately calculate census data, the increase in unemployment could also be the result of census taking methods. The article concludes that socio-economic data should be available at the smallest possible geographic area to ensure more accurate results in interpolation. It also recommends that new output areas be conceptualised to create a seamless database of census data from 1991 to 2011 in South Africa.

Introduction

One of the main interests of census and population researchers is the study of socio-economic change.[1] The study of change is especially important to answer questions at a local scale, for example, a suburb or municipality. However, such comparisons over time are difficult because census collection methods and definitions as well as reporting geographies change.

This article seeks to find a solution to the problem of shifting boundaries as exemplified in the case of unemployment trends over time in South Africa. The problem of shifting boundaries is referred to as the modifiable areal unit problem (MAUP). Currently there is limited literature available on overcoming the MAUP in South Africa using socio-economic data. Although there are some attempts to build historical Geographic Information System socio-economic data sets in South Africa, their methodologies are not yet documented in the literature.[2]

To minimise the effects of the MAUP, administrative units should be as disaggregated as possible.[3] The MAUP is composed of two problems: firstly, the scale problem, where different results can occur when one set of areal units is aggregated into a fewer number of larger units for analysis; and secondly, the aggregation problem, where different results can be obtained when boundaries of spatial entities are arranged in different ways. In this article, the MAUP specifically refers to the change in geographical units of analysis.

Given the challenge of enumeration area (EA) and magisterial district boundary changes in South Africa since 1991, this article seeks to find a methodological solution for overcoming modifiable areal unit problems in South African census data. Unemployment, which is extremely high in South Africa, and often used as an indication of socio-economic well-being, is used here to illustrate the challenge of the MAUP.

Challenges in integrating spatial and temporal data over time

The historical linking of data in Geographic Information Systems (GIS) faces two major problems. Firstly, available commercial software is ill suited to temporal GIS. Secondly, and partly in consequence, historical GIS construction is very expensive, especially because of labour costs related to extensively solving challenges.[4] These factors influence any longitudinal spatial analysis of socio-economic conditions in South Africa.

Census data are the only data at a national level that include all citizens and cover the full geographical extent of the country. Therefore these data were accepted as the spatially most comprehensive data on unemployment in South Africa. There are, however, arguments that the census is not as accurate as the Labour Force Survey (LFS) and Quarterly Labour Force Survey (QLFS) in measuring unemployment because it does not ask questions about the reasons for not being employed. Nevertheless, it was important to use the census, which provides the most spatially detailed data in order to build a national understanding of an issue at a spatially disaggregated (sub-provincial) level.

The study of change over time through reference to census data is fraught with difficulties because of operational changes made between censuses to improve their relevance and reliability, and because of differences in the degree to which they achieved their aims.[5] Although aggregation to geographical areas is a near-universal feature of census information, it is fundamentally difficult to accommodate.[6] Data created by interpolation are estimates that will inevitably contain a certain degree of error.[7] The accuracy of areal interpolation, moreover, will vary according to the nature of the variable being interpolated, the nature of the ancillary data and the shape and size of both the source and target units.[7]

Further challenges in terms of integrating spatial data over time are the scale of administrative units for analysis and the linking of the attribute data from various years to the spatial data.[8,9] Although these two aspects are related, they pose distinct ways of dealing with the problem. The remainder of this section deals with these two issues separately using unemployment data to illustrate them.

Means of addressing the modifiable areal unit problems

Spatial solutions

One of the ways to link different census geographies is through areal interpolation. Areal interpolation was used to create a temporal census database for Europe and areal weighting specifically was used to achieve this.[4] Areal weighting is based on the assumption that data are distributed homogeneously across each source unit. Count data for a variable y are interpolated from the source zones to the target zones using the formula:

$$\hat{y}t = \sum [Ast / As \times ys] \qquad \text{Equation 1}$$

where $\hat{y}t$ is the estimated value for the target zone, ys is value for the source zone, As is the area of the source zone, and Ast is the area of the zone of intersection between the source and target zones.[10] The assumption of a homogeneous population distribution is, however, unrealistic for most socio-economic applications. Although a number of studies have tried to render this mathematical model more flexible by introducing ancillary data to the model, no matter how good the technique, areal interpolation will inevitably contain some error, the impact of which will vary from polygon to polygon.[5]

The nature of the data being interpolated also determines the accuracy of the results.[11] Some attempts have been made to calculate the interpolation error, but most of these are limited because they show global goodness of fit and do not provide detail at the level of individual data values. In statistical terms 'goodness of fit' refers to the confidence with which a model can be presented while 'global' indicates the fit of that model at the general level. Poorer results are obtained as the scale of the output area is reduced and more accurate interpolation is achieved in rural areas.[12] Urban areas are complex and there seems to be no easy solution to the challenge. The choice of an interpolation strategy has a strong influence on model results, and thus on potentially far-reaching policy decisions.[13] It is therefore important to use the correct methodology.

Three types of re-aggregation criteria are offered to create new areal units from census data: firstly, areas that possess similar levels of heterogeneity for the specific variable of interest; secondly, areas that are of an approximately equal size and shape; and thirdly, areas that provide an efficient partitioning of space and are of similar nature.[14]

Besides areal interpolation, small area grids and automated zoning can be used. These methods are briefly discussed and some examples given. The centroids of the1991 enumeration districts in England, Wales and Scotland were used in an areal aggregation solution to aggregate data to the 1981 wards.[1] The same was done for the 1971 enumeration district data, whereafter it was aggregated to higher geographies. A controlled public access system was created, which delivers a number of user defined outputs.

In a different technique, root mean square (RMS) error is used to quantify the error introduced in simulations. This is based on the average differences between the estimated values of the variable and its known actual values.[15]

Another solution to represent data from small areas in a re-aggregated format is the use of grids. Small area grids should be more than 1 km (5 km is desired).[16] On the other hand, small grids have the risk of containing below-threshold numbers of people thereby compromising anonymity. A greater understanding of the uses to which grid-based geographies are put is required in order to assess whether such grids would prove useful.[16] Research showed that polygon methods have a lower map accuracy error than their grid-based counterparts, though the difference was not statistically significant.[17]

The most appropriate response to the MAUP is to design purpose-specific zonal systems.[3] A study using such methods aggregated data from the household level to small areas (data zones) using the Community Survey (CS) 2007 data in South Africa.[18] The authors used multi-level modelling to calculate a 'best linear unbiased estimator' of multiple deprivation for each data zone in the country. The data zones were calculated based on a data zone code provided by Statistics South Africa (Stats SA) and multi-level modelling was used for attribute data.

Continual boundary revision of census areas in order to retain a degree of equality of population sizes renders the separation of population change from boundary change particularly difficult.[19] It would therefore become difficult to detect real population change, because such change would always be associated with a change in boundaries.

Temporal solutions

Stats SA produces all the official data sets for South Africa. These include censuses, LFS and CS, among others. Censuses take place every 5–10 years and cover the total population, while the QLFS takes place quarterly and relies on a national sample of about 30 000 households.[20] The QLFS reports unemployment figures at a provincial level and distinguishes between metropolitan and non-metropolitan areas within a province, while censuses report unemployment data at an EA, magisterial district or municipality level. In the absence of a census in 2006, a CS was conducted in 2007 and reported data at a municipal level.

Table 1 summarises the reporting geographies of Stats SA products and shows that it varies significantly between different products. Census, which contains the most detailed spatial data, takes place every 5 years or alternatively every 10 years. Labour Force Surveys only report unemployment at a provincial level, i.e. one record for every province.

Table 1: Summary of reporting geographies of various Stats SA products

Data source	Reporting geography	Interval between reporting periods
Census	Enumeration Area (1991, 1996) Sub-place (2001) Small Area Layer (2011) Magisterial district (1991, 1996) Municipality (2001, 2011)	Supposedly 5 years, alternatively 10 years
Community Survey	Municipality	Once off
Labour Force Survey	Province (2001–2007)	6 months
Quarterly Labour Force Survey	Province and metro for those provinces with metros (2008+)	Quarterly

Stats SA has admitted that the following factors complicate the use of comparative studies that rely on data collected in South African censuses.[21] Firstly, the continuous and complete changing of administrative boundaries; secondly, the revision of the set of EAs that was used in Census 1996; and thirdly, the decision not to release the Census 2001 data at EA level. The fact that spatial units from the various censuses and surveys are not the same at different periods creates problems for researchers attempting to establish a seamless, temporal database for analysis of unemployment.

The effect of census change in South Africa

EAs are essentially used to make the work of the census enumerator manageable, so these boundaries change between censuses. The number of EAs has increased significantly since 1991 to accommodate changes in population growth, urbanisation and so forth. Most provinces had a moderate decline in the number of EAs between 1996 and 2001 while the total number of EAs in the 2011 census increased to around 104 000.[22]

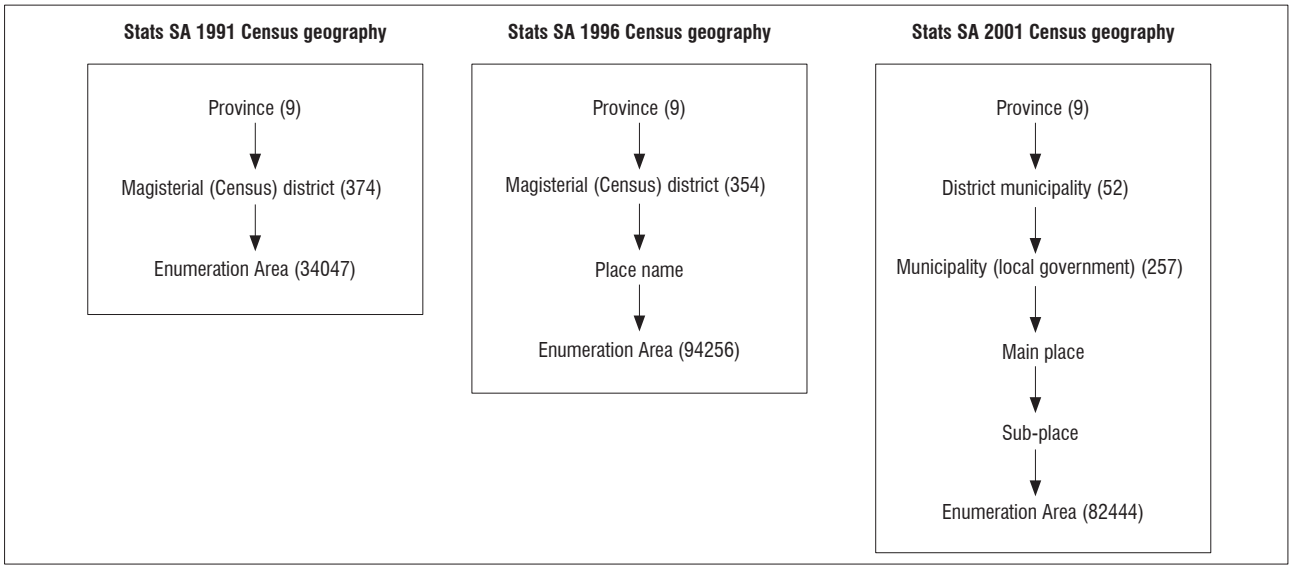


Figure 1: Spatial hierarchy of various censuses.

Figure 1 illustrates the differences between the spatial units in the various censuses. In 1991, census data were gathered at an EA level and disseminated at a magisterial district level. In 1996, the situation was the same except that data were released at an EA level as well. In 2001, Stats SA created the first official census geography with a number of spatial levels; and although the data were collected at an EA level, the information was only released at a sub-place level.

In order to obtain a trustworthy understanding of spatial changes in unemployment patterns over a given time span, data from four different periods are required. Data from only three periods might possibly not show a trustworthy trend in unemployment. The most recent censuses in South Africa were conducted in 2011, 2001, 1996, 1991, 1985, 1980 and 1970.

The censuses of 1985 and 1980 did not include all South Africans and were therefore problematic for this research. Although the census of 1970 included South Africans of all race groups, labour patterns have changed significantly since then and would not be relevant to current, post-apartheid South Africa. Therefore, another data source was considered, the CS 2007, which was a large sample survey that presented results at the level of the local municipal boundaries of 2005.[23]

Methodology

The challenge of analysing unemployment spatially over time is the incompatibility of various spatial units of data representation. Ideally, areal interpolation should be done at the most detailed level spatially and in the case of South Africa, this would be the EA level. As the EA is the smallest geographical unit, it should make aggregation to higher-level geographical units easy and achievable. However, the difficulty in matching attribute and spatial data (1996 census) and the lack of attribute data (1991 census) makes this difficult to attain.

For this research, the linking of different census geographies was achieved by using areal interpolation to transfer data from one set of boundaries to another. The 2005 municipality boundaries were used as the common denominator – part of a spatial hierarchy developed by Stats SA for the 2001 census. This hierarchy started with the EA as the lowest building block, then sub-place, main place, local municipality and province (see Figure 1). The municipal boundaries for 2005 were chosen as the preferred target area, because they represented a recent administrative division useful for displaying temporal change and they were easily linked to the 2001 and 2007 data.

Other methodologies that could have been considered were, for example, disaggregation based on dasymetric mapping but this method required additional verification through ancillary data.[2,17,24] The lack of an additional data set related to economic activities at municipal level would makes the validation process difficult.[25]

Data sources

Census data constitute the only data source at a national level that aims to include all citizens and all geographical areas. It was therefore accepted that the census was the most comprehensive spatial data set on unemployment in the country. Census and CS data were obtained from Stats SA and the Human Sciences Research Council (HSRC) of South Africa.

In the 1991 census, unemployment statistics data were not directly calculated; these figures were generated by subtracting the number of employed people from the economically active population. The estimated undercount in 1991 was an average of 12.7% across all race groups.[26] Stats SA provided the number of unemployed people per magisterial district in the 1996 census. The statistics for these two censuses were therefore in comparable formats. The estimated undercount in 1996 was 10.7%.[27]

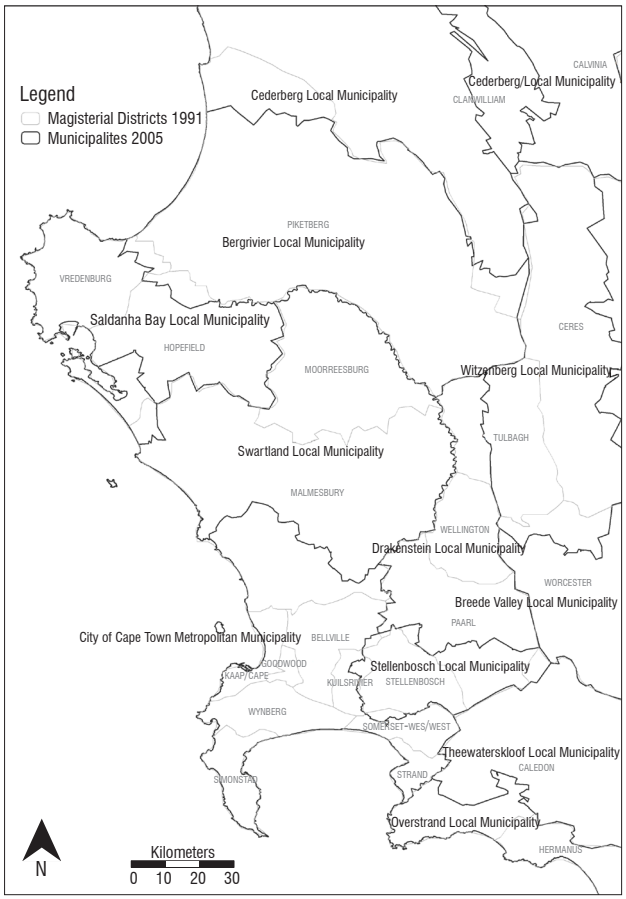
The 2001 census attribute data were not released at an EA level for reasons of confidentiality[21] but were made available at a sub-place level and could therefore be aggregated to municipal level. The estimated undercount for people in the 2001 census was 17.6%.[28] CS 2007 released data on the number of unemployed people as well as the economically active population at municipal level. The level of analysis was therefore standardised to municipalities and depended on a certain measure of areal interpolation – the most optimal way of standardising the unit of analysis.[8]

It was decided that the CS 2007 rather than the 1970 Census would be used because the CS was closer in time to the other censuses used. Although the CS 2007 was a survey, the sample size of 949 105 persons was large enough to enable reporting of the findings at municipal level.

The correction of the undercount in census 1996 and census 2001 was undertaken during the post-enumeration survey. After the 1991 census, an adjustment for the undercount was made and the corrected statistics were released. In the CS 2007, the estimation process was based on the ratio method of projecting geographic subdivisions to determine the populations of district councils and municipalities. This article did not investigate the undercount corrections per se, and assumed that the data released by Stats SA were accurate and thus constituted the most complete data available spatially.

Aggregating data

Areal aggregation is usually from a source polygon layer to a target polygon layer. Figure 2 shows a layout of the magisterial and municipal boundaries in Western Cape and displays that magisterial districts were largely within the boundaries of metropolitan areas. In non-metropolitan municipalities, magisterial boundaries sometimes extended beyond those of the municipality. Magisterial districts were used as the source polygon and municipalities as the target polygon.



Sources: Department of Justice 1991[29] and Municipal Demarcation Board 2005[30]

Figure 2: Example of magisterial district and municipal boundaries in Western Cape

A geostatistical process of areal-based interpolation was followed and it initially created a prediction surface from the source polygons. Predictions and standard errors were calculated for all points within and between the input polygons (magisterial districts). As unemployment statistics are continuous, the input was defined as Gaussian polygonal because the polygons of collection were fairly large.[31] A standard error of the predicted unemployment value was calculated for each polygon and is further discussed in conjuction with Figure 5.

During the second part of the interpolation, a process of cross-validation and validation was followed to make a decision as to which model provided the best predictions. These predictions, along with standard errors, were calculated for each municipal polygon. Some of the other components of this variogram stage included model type, lag size and lattice spacing (Figure 3). The variogram model can either be circular, spherical, exponential, tetraspherical, pentaspherical, Gaussian, K-Bessel, Stable and J-Bessel. The spherical model is the most widely used variogram model and assumes that covariance reaches a value of zero at a specific distance.[32] K-Bessel and Stable models produce the best results because they take an additional shape parameter that allows the model to change curvature.[33] However, they also take the longest to process.

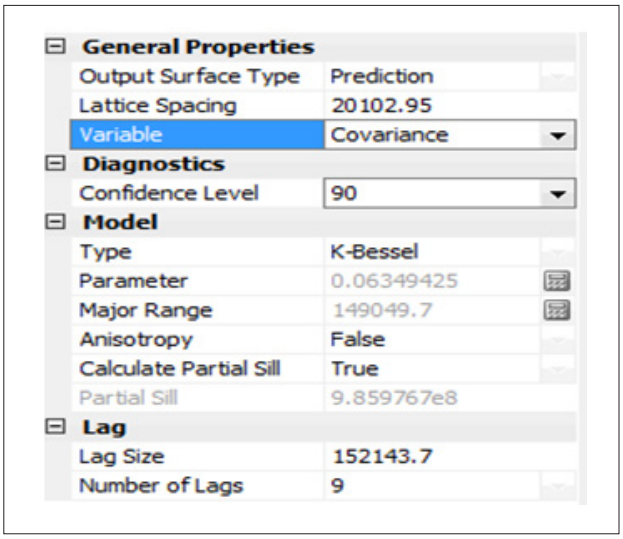


Figure 3: Components of the variogram stage of areal interpolation adapted from ArcGIS.[33]

Lag size refers to the number of adjacent cells in a straight horizontal or straight vertical line from the centre to the edge of the figure and usually the lag size multiplied by the number of lags should be less than one-half of the largest distances in one's data set.[31] Fewer lags usually improve the covariance value. Lattice spacing refers to the number of points used within each polygon to build the semi-variogram and the smaller the spacing, the longer it takes to generate the semi-variogram and the more memory is consumed.

For a model to provide accurate predictions, the standardised mean error should be close to 0, the root-mean-square error and average standard error should be as small as possible and the root-mean square standardised error should be close to 1.[31]

1991 and 1996 Census

For the 1991 data, a spherical model was used and twelve lags were performed within a confidence level of 90%. The final model of the 1996 data used a K-Bessel model with nine lags, a lag size of 100 000 and a lattice spacing of 18 000. The K-Bessel model is more stable, but also takes longer to process. If the models were correctly specified, one would expect 90% of the empirical co-variances to fall within the confidence intervals.

2001 Census

To obtain unemployment figures for 2001, data were downloaded from the 'Statistics SA Interactive data' website (http://interactive.statssa.gov.za:8282/webview/). Because these data were pre-customised for the 2005 municipal boundaries, there was no need to perform any calculations on the data to re-align them with the 2005 boundaries.

2007 Community survey

The 2007 CS data were downloaded from the 'Statistics SA Interactive data' website. The data reported results at the spatial level of the 2005 municipal boundaries. The unemployment rate was then calculated using the following formula (which excluded institutions like hostels, retirement homes, etc. and 'unspecified' responses):

$$[\text{unemployed}/(\text{employed} + \text{unemployed})] \times 100 \qquad \text{Equation 2}$$

Data for each year being analysed had to be treated individually because there was no global solution to the spatial problems encountered. The unemployment definitions remained the same and were therefore comparable. Data pre-processing was time-consuming, which confirmed the assertion that temporal construction in GIS is a costly exercise.[34] In this case the data were obtained free of charge, but it took a considerable amount of time to process the data to a comparable format.

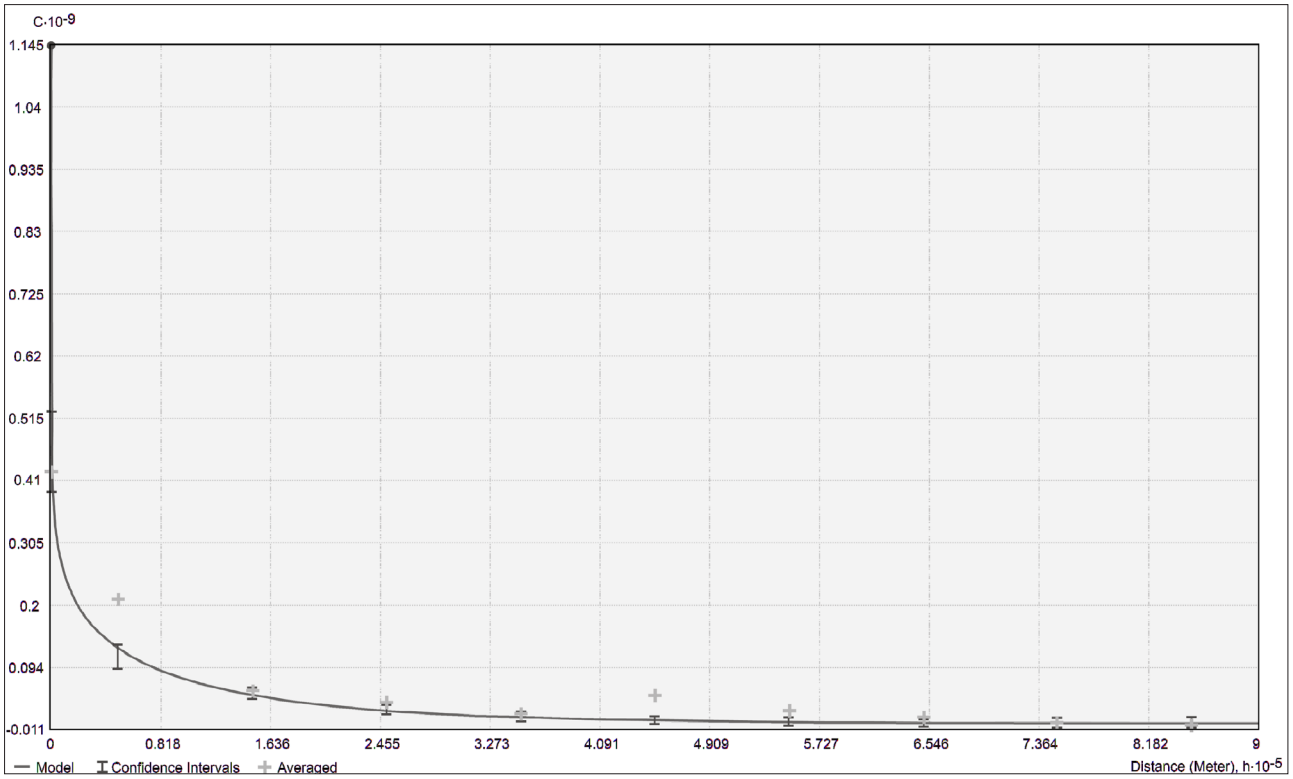


Figure 4: Variogram of the 1996 unemployment data.

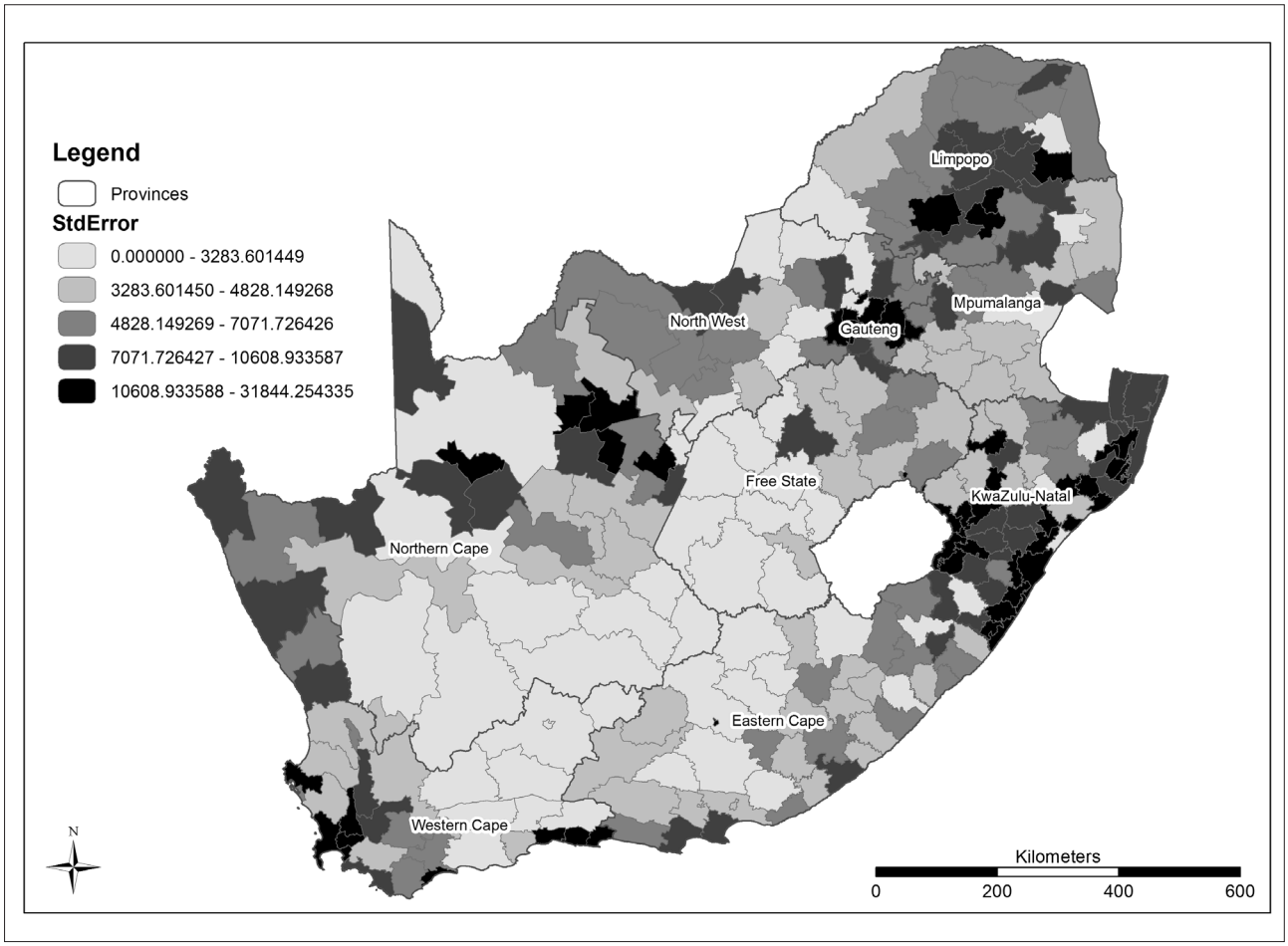


Figure 5: Standard errors in the 1996 predicted unemployment data.

Findings and discussion

Results from the 1991 areal interpolation show that 8 of the 13 points were within the 90% confidence level so the model is therefore acceptable. Furthermore, the root mean squared standardised error was 1.2, therefore the predicted standard errors were valid because they are close to 1.

Figure 4 displays the variogram with averaged values (thick crosses) for the 1996 data and more than 90% of the empirical co-variances fell within the confidence intervals (long bars). The model was therefore acceptable. The mean squared standardised error was 0.9 and therefore the predicted standard errors were valid.

Figure 5 displays the standard errors for the 1996 data in a quantile map for the target layer, namely the 2005 municipalities. The largest errors were recorded in metropolitan areas like Cape Town, Johannesburg, Pretoria, Ekurhuleni and eThekwini. However, the large errors were not restricted to metropolitan municipalities and also appeared in other urban municipalities for example along the coast of KwaZulu-Natal, around Polokwane in Limpopo, the southern coast and selected municipalities in the Western Cape. This finding strengthens literature trends that established more accurate interpolation results in rural areas.

Table 2 indicates that in 1991, 5 of the 257 municipalities had more than 1000 EAs. All these municipalities were metropolitan areas. In 1996, 11 of the 257 municipalities (4.2%) had more than 1000 EAs. Besides the metropolitan areas, a number of municipalities in the Eastern Cape (including the former Transkei) and in Gauteng and KwaZulu-Natal had more than 1000 EAs each.

The implication is that in 1996 in 4.2% of municipalities, the aggregation error could be bigger than the census undercount. For the 1991 data, this would be 2% of municipalities. Therefore, these simple aggregation techniques would result in an aggregation error in a small percentage of the data which could be considered acceptable.[1]

Results from the areal interpolation of the 1991 and 1996 magisterial districts used in this article yielded the statistics shown in Table 3. Selected municipalities are reflected in the table as there are too many municipalities to portray at once. The table shows results for two municipalities in each of the nine provinces, representing different types of entities – urban/metropolitan and rural. The data in the table are sorted alphabetically by province.

The municipality discussed here with the highest unemployment rate was Nongoma, in KwaZulu-Natal. This municipality showed continuous high levels of unemployment in 1996, 2001 and 2007. There was an increase of about 20% in unemployment between 1996 and 2001, while the figures declined somewhat in 2007.

Large increases were seen between the two sets of 1996 and 2001 statistics, especially in the metros, for example, the City of Cape Town that recorded 6.9% unemployment in 1996. In metros like Cape Town, Johannesburg and eThekwini, the same differences prevailed in 2001 and 2007. The inflated unemployment rates for 2001 can be ascribed to misclassification, where employment status shifted some of discouraged job seekers to outside the economically active band.[25]

Sparsely populated areas like Matatiele (Eastern Cape) and Kamiesberg (Northern Cape) also showed stark differences in the percentages of unemployed people. The official statistics (that is, those sourced from Stats SA) show peak unemployment figures in 2001 with moderate declines in 2007.

Considering the different ways of overcoming the census MAUP in South Africa, the ideal solution would be to use the boundaries of existing small area features like EAs or sub-places. The challenge of not having attribute data for all EAs in 1991 and 1996 made this difficult to attain. The same shortcoming would apply to small area grids and automated zoning, because the lack of underlying data would make it difficult to achieve results at small area levels. Automated zoning could be used at higher level geographies like municipalities, but as data existed for magisterial districts in 1991 and 1996, it was more opportune to aggregate these to the 2005 municipalities.

A limitation of the uncertainty estimates provided by areal interpolation methods is that they are likely to underestimate the real uncertainty of the associated spatial predictions. This is based on the fact that these methods are two-stage processes, where a variogram model is estimated in the first stage and prediction and uncertainty equations are applied in the second stage. The second stage equations treat the first stage variogram model as true and known.[37] To address this, the documentation of any spatial interpolation exercise can explicitly recognise their bias toward underestimating uncertainty. Decision makers and other users of the modelling results can then decide for themselves how best to adjust their interpretation based on this information.[37]

Kriging-based methods such as areal interpolation are by definition smoothing operators. Although the formulas are true to the data, the relatively smooth spatial interpolation surface they provide may not capture all the underlying variation of the process they are attempting to describe.[37] In this research a discreet input and output value was calculated for unemployment in a specific municipality. The value cannot be reflective of all unemployment rates across that municipality, e.g. unemployment rates in a township is not the same as unemployment in a suburb or the city centre. In fact, to require an interpolation model to be highly accurate at each specific location is an unfair expectation, because it is supposed to provide a spatial description of the unemployment situation in general.[37]

Table 2: Number of enumeration areas (EAs) in municipalities with more than 1000 EAs

Municipality	Province	EAs in 1991	EAs in 1996
Buffalo City	Eastern Cape	379	1767
Cape Town Metro	Western Cape	2950	4760
Ekurhuleni Metro	Gauteng	1344	5027
Emfuleni	Gauteng	340	1404
Ethekwini (Durban Metro)	KwaZulu-Natal	2624	6414
Johannesburg Metro	Gauteng	2215	6148
King Sabata Dalindyebo	Eastern Cape	2	1399
Mbhashe	Eastern Cape	3	1081
Mnquma	Eastern Cape	3	1085
Msunduzi	KwaZulu-Natal	435	1031
Tshwane Metro	Gauteng	2107	3595

Table 3: Unemployment rates in selected municipalities

Municipality name	Province	Author's calculations 1991	Author's calculations 1996	Stats SA 2001[35]	Stats SA 2007[36]
Matatiele	Eastern Cape	17.2	28.8	62.7	38.6
Nelson Mandela Bay Metro	Eastern Cape	25.8	8.0	46.4	40.0
Mangaung	Free State	23.7	9.8	40.1	29.9
Maluti a Phofung	Free State	32.6	20.2	57.5	51.4
Mogale City	Gauteng	16.5	6.1	34.1	25.7
City of Johannesburg Metro	Gauteng	20.3	6.7	37.3	29.6
Nongoma	KwaZulu-Natal	52.2	51.8	71.7	63.2
eThekwini Metro	KwaZulu-Natal	23.2	10.1	43.0	33.4
Mkhondo	Mpumalanga	12.7	10.4	45.8	43.3
Mbombela	Mpumalanga	22.6	9.9	37.8	24.0
Kamiesberg	Northern Cape	15.8	15.5	32.0	28.4
Sol Plaatjie	Northern Cape	24.0	9.2	41.5	32.7
Musina	Limpopo	15.8	7.5	24.9	19.6
Polokwane	Limpopo	27.9	22.7	41.5	37.2
Mafikeng	North West	32.6	19.4	49.3	42.8
City of Matlosana	North West	12.7	7.2	40.0	31.5
City of Cape Town Metro	Western Cape	18.1	6.9	29.2	24.5
Beaufort West	Western Cape	21.1	11.3	39.1	32.0

Conclusion and recommendations

Empirical analysis of past trends is vital for extending knowledge of the processes producing change. This article aimed to create a spatially comparable unemployment data set from 1991 to 2007. Although there are enormous challenges related to constructing a time-continuous GIS data set, many of these were overcome by aggregating data from magisterial district boundaries to municipalities. The research has created a spatio-temporal data set which could be used as a basis for future research and can be regarded as work in progress.

Accuracy was reflected in prediction errors for each polygon and this lends validity to the findings. The predicted standard errors for the 1996 data were valid, because the mean squared standardised error was 0.9 which is close to the acceptable value of 1. Similarly, the 1991 results were acceptable because the root mean standardised error was close to 1.

Inaccuracies in the 1991 and 1996 EA level data made it difficult to accurately aggregate to higher geographical entities. The compromise was to aggregate data from magisterial districts for these years to municipal boundaries. Future research on the South African data could include calculations such as weighting the interpolation process, which might increase the confidence with which one could report reasonable results. Weighting could be done using population density or road density. However, if unemployment statistics are released at a local scale such as sub-place, weighting would not be necessary, because the variance within a polygon should be minimal.

The areal interpolation results indicated that the largest errors were recorded in metropolitan areas and other urban municipalities. This finding supports literature trends which established that more accurate interpolation tends to be in rural areas. Further research could investigate the possibility of using small grids as an alternative method to aggregate unemployment data and compare results with the findings of this research.

The best way to overcome some of the errors in the findings is to ground-truth the data. This would include physically counting how many people were unemployed in a specific area. As the data date from 1991 and 1996, the only proxy for such a count would be to use satellite imagery of that time, count the number of dwellings and then estimate the number of unemployed based on a ratio of employed people vs unemployed people. To ground-truth data in the field is a very costly exercise and that is one of the reasons why a census is not conducted every year.

A lack of spatially detailed data is at the helm of this research problem and other methods like dasymetric mapping, which is dependent on supplementary data at a detailed spatial level, will suffer the same shortcomings. It is recommended that aggregated data on unemployment and other socio-economic variables be created from the smallest spatial unit, that is, EA or sub-place. However, in the South African case, EA boundaries changed again with Census 2011, which means that researchers would have to once again recalculate socio-economic data to fit the new features.

A further possible solution could be to impute unemployment data for 1991 and 1996 EAs in cases where such data are missing. This would allow the aggregation of unemployment and other census data to larger geographical units. Data incompatibilities could also be overcome by creating new output areas that would be able to serve as target areas for areal interpolation from the 1991 census onwards. Alternatively, Census 2011 EA centroids could be used to interpolate attribute data from EAs of earlier years.

By highlighting and addressing some of the spatial and attribute data challenges of publicly available unemployment data in South Africa,

this article has created a base for future research using the same data sources. The research aimed to create a comparative geographical database where socio-economic change is not the result of boundary changes. This article has shown that there are still a number of hurdles to overcome in creating a seamless database of census data since 1991. Hopefully the results from Census 2011 will allow the creation of a longitudinal data set of spatial socio-economic trends in South Africa at a detailed geographical level.

Acknowledgements

The author would like to thank Adlai Davids and Fethi Ahmed for valuable input received on earlier versions of this article.

References

1. Martin D, Dorling D, Mitchell R. Linking censuses through time: Problems and solutions. Area. 2002;34:82–91. http://dx.doi.org/10.1111/1475-4762.00059
2. Mans G. Developing a geo-data frame using dasymetric mapping principles to facilitate data integration [document on the Internet]. c2010 [cited 2012 April 06]. Available from: http://www.gap.csir.co.za/gap/gap-applications
3. Openshaw S. The modifiable area unit problem. Concepts and techniques in modern geography 38. Norwich: Geo Books; 1984.
4. Gregory IN, Southall H. Spatial frameworks for historical censuses: The Great Britain historical GIS. In: Hall PK, McCaa, R, Thorvaldsen G, editors. Handbook of historical microdata for population research. Minneapolis: Minnesota Population Center; 2000. p. 319–33.
5. Champion AG. Analysis of change through time. In: Openshaw S, editor. Census users' handbook. Cambridge: GeoInformation International; 1995. p. 307–335.
6. Martin D. 2000. Census 2001: Making the best of zonal geographies. Paper delivered at The Census of Population: 2000 and Beyond; 2001 June 23; Manchester, England.
7. Gregory IN, Ell PS. Error-sensitive historical GIS: Identifying areal interpolation errors in time-series data. Int J Geogr Inf Sci. 2006;20(2):135–152. http://dx.doi.org/10.1080/13658810500399589
8. Gregory IN, Marti-Henneberg J, Tapiador F. A GIS reconstruction of the population of Europe, 1870 to 2000 [document on the Internet]. c2008 [cited 2011 Jul 19]. Available from: http://web.udl.es/dept/geosoc/europa/cas/img/GIS%20Approaches%20-%20resubmission.6.06.pdf
9. Martin D, Gascoigne R. Change and change again: Geographical implications for intercensal analysis. Area. 1994;26:133–141.
10. Goodchild MF, Lam NSN. Areal interpolation: A variant of the traditional problem. Geo-Processing. 1980;1:297–312.
11. Gregory IN. The accuracy of areal interpolation techniques: Standardising 19th and 20th century census data to allow long-term comparisons. Comput Environ Urban. 2002;26:293–314. http://dx.doi.org/10.1016/s0198-9715(01)00013-8
12. Martin D. Extending the automated zoning procedure to reconcile incompatible zoning systems. Int J Geogr Inf Sci. 2003;17:181–196. http://dx.doi.org/10.1080/713811750
13. Goodchild MF, Anselin L, Deichmann U. A framework for the areal interpolation of socioeconomic data. Environ Plan. 1993;25:383–397. http://dx.doi.org/10.1068/a250383
14. Charlton M, Rao L, Carver S. GIS and the census. In: Openshaw S, editor. Census users' handbook. Cambridge: GeoInformation International; 1995. p. 133–166. http://dx.doi.org/10.1068/a270425
15. Fisher PF, Langford M. Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. Environ Plan. 1995;27:211–224. http://dx.doi.org/10.1068/a270211
16. Duke-Williams O, Rees P. Can census offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure. Int J Geogr Inf Sci. 1998;12:579–605. http://dx.doi.org/10.1080/136588198241680
17. Eicher CL, Brewer CA. Dasymetric mapping and areal interpolation: Implementation and evaluation. Cartogr Geogr Inf Sci. 2001;28:125–138. http://dx.doi.org/10.1559/152304001782173727
18. Noble M, Dibben C, Wright G. The South African index of multiple deprivation 2007 at datazone level (modelled). Pretoria: Department of Social Development; 2010.
19. Geddes A, Flowerdew R. Geographical considerations in designing policy-relevant regions. Proceedings of the 3rd AGILE Conference on Geographic Information Science; 2000 May 25–27; Helsinki: Finland. p. 80–81
20. Statistics South Africa (Stats SA). Guide to the Quarterly Labour Force Survey August 2008. Pretoria: Stats SA; 2011.
21. Statistics South Africa. Using the 2001 Census: Approaches to analysing data [document on the Internet]. c2007 [cited 2010 Aug 22]. Available from: http://www.statssa.gov.za/Publications/CensusHandBook/CensusHandbook.pdf
22. Statistics South Africa (Stats SA). Census 2011: EA spatial data. Data set: Pretoria: Stats SA; 2012.
23. Statistics South Africa (Stats SA). General Household Survey: Report no. P0318. Pretoria: Stats SA; 2008.
24. Mennis J. Generating surface models of population using dasymetric mapping. Prof Geogr. 2003;55(1):31–42.
25. Hakizimana JV. Small area estimation of unemployment for South African labour market statistics [Msc dissertation]. Johannesburg: University of Witwatersrand; 2011.
26. Statistics South Africa. Census 1991: Explanatory notes [document on the Internet]. no date [cited 2011 Apr 15]. Available from: http://interactive.statssa.gov.za:8282/metadata/censuses/1991/RSA/RSA1991.htm
27. Statistics South Africa (Stats SA). Calculating the undercount in Census '96. Pretoria: Stats SA; 1998.
28. Statistics South Africa (Stats SA). Census 2001: How the count was done. Pretoria: Stats SA; 2003.
29. Department of Justice. Magisterial districts: Dataset. Pretoria: Department of Justice; 1991.
30. Municipal Demarcation Board. Municipal boundaries. Dataset. [Information on the Internet]. c2005 [cited 2005 Jul 15]. Available from: http://www.demarcation.org.za/
31. Krivoruchko K, Gribov A, Krause E. Multivariate areal interpolation for continuous and count data. Procedia Environ Sci. 2011;3:14–19. http://dx.doi.org/10.1016/j.proenv.2011.02.004
32. Brusilovskiy E. Spatial interpolation: A brief introduction. Business Intelligence Solutions [article on the Internet]. no date [cited 2015 June 24]. Available from: http://www.bisolutions.us/A-Brief-Introduction-to-Spatial-Interpolation/php
33. Johnston K, Ver hoef JMV, Krivoruchko K, Lucas N. Using ArcGIS Geostatistical Analyst (version 9). Redlands: Environmental Systems Research Institute; 2003.
34. Gregory IN and Southall H. Spatial frameworks for historical censuses: The Great Britain historical GIS. IPUMS. 2003. Available from: https://international.ipums.org/international/resources/microdata_handbook/2_03_spatial_frameworks_ch19.pdf
35. Statistics South Africa (Stats SA). Census 2001: Dataset. Pretoria: Stats SA; 2001.
36. Statistics South Africa (Stats SA). Community survey: Dataset. Pretoria: Stats SA; 2007.
37. Environmental Protection Agency (EPA). 2004. Developing spatially interpolated surfaces and estimating uncertainty. EPA-454/R-04-004. Durham: EPA.

Dental microwear differences between eastern and southern African fossil bovids and hominins

AUTHORS:
Peter S. Ungar[1]
Jessica R. Scott[2]
Christine M. Steininger[3,4]

AFFILIATIONS:
[1]Department of Anthropology, University of Arkansas, Fayetteville, Arkansas, USA
[2]Department of Sociology and Anthropology, University of Arkansas, Little Rock, Arkansas, USA
[3]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa
[4]Centre of Excellence in Palaeosciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Peter Ungar

EMAIL:
pungar@uark.edu

POSTAL ADDRESS:
Department of Anthropology, University of Arkansas, Old Main 330, Fayetteville, Arkansas 72701, USA

DATES:
Received: 11 Oct. 2015
Revised: 18 Dec. 2015
Accepted: 28 Dec. 2015

KEYWORDS:
grit; diet; habitat; fossil ruminants; tooth wear

HOW TO CITE:
Ungar PS, Scott JR, Steininger CM. Dental microwear differences between eastern and southern African fossil bovids and hominins. S Afr J Sci. 2016;112(3/4), Art. #2015-0393, 5 pages. http://dx.doi.org/10.17159/sajs.2016/20150393

Dental microwear has proven to be a valuable tool for reconstructing diets of fossil vertebrates. However, recent studies have suggested that the pattern of microscopic scratches and pits on teeth may be more reflective of environmental grit than of food preferences. Could differences in dental microwear between early hominins, for example, therefore be a result of dust level rather than of diet? We investigated this possibility using a palaeocommunity approach. We compared microwear texture differences between eastern and southern African Hominini, along with Plio-Pleistocene specimens representing two tribes of bovids, Alcelaphini and Antilopini, from the same deposits as the early hominins. If exogenous grit swamps diet signals, we would expect community-wide microwear patterns separating samples by region. Results indicate that each of the three tribes shows a different pattern of variation of microwear textures between eastern and southern Africa. These results imply that differences in microwear reflect diet rather than grit load, and that microwear can provide valuable information not just about environmental dust level, but about food preferences of fossil vertebrates.

Dental microwear is an important tool for reconstructing diets of fossil vertebrates, from Palaeozoic conodonts[1] to Plio-Pleistocene hominins[2]. Microwear researchers have noted consistent and predictable relationships between pattern and behaviour in extant taxa from fishes[3] to humans[4]; and these relationships have been used as a baseline to infer the diet of extinct species from their teeth. The basic assumption for mammalian cheek teeth has been that hard foods are crushed, causing pitting as opposing surfaces are pressed together, whereas tough foods are sheared, causing scratches as abrasives are drawn along opposing surfaces that slide past one another.[5] For example, primates that eat hard nuts and palm fronds tend to have more microscopic dental pits than primates that eat tough leaves.[6] This diet–microwear pattern association has been used to infer feeding behaviours of many fossil species, including early hominins.[7]

However, a recent study[8] has called into question the efficacy of microwear as a proxy for diet, suggesting that experimental validation is needed to affirm relationships between pattern and foods eaten. For example, in an in vitro wear simulation study, Lucas and coauthors[9] found that while quartz dust on foods can easily wear tooth enamel, phytoliths within them might not. This finding has led some to suggest that grit in the environment may be more important to wear pattern than factors intrinsic to items eaten.[10] In fact, Strait et al.[8] argued that microwear patterns for early hominins may reflect the dustiness of the environment and, 'say little about the nature of the foods themselves'[8(p.348)]. The argument follows that the more striated and less pitted microwear seen for Plio-Pleistocene hominins from eastern Africa than those from southern Africa[7] may have more to do with where they lived than what they ate. This possibility has important implications not only for studies of early hominins, but also for the countless other fossil vertebrates for which microwear has been documented and related to diet.[11]

Strait et al.[8] suggested that quartz dust might cause heavy microwear pitting, and Williams[12] opined that exogenous grit could lead to especially complex surface textures. These suggestions do not explain the lack of such pitting on *Paranthropus boisei* teeth, which have been suggested to evince extreme macrowear indicative of a gritty, abrasive environment.[10] But they do raise the question: Is there a consistent relationship between environmental-grit level and microwear? While it is clear that soil quartz levels can play an important role in tooth wear[13,14], studies of mammals living in different settings today have failed to find that grit obscures diet-related microwear signals[15].

But what about the differences between Plio-Pleistocene hominins from eastern and southern Africa? If diet signals are 'swamped' by grit, it should be the case not only for early hominins, but also for other taxa. We predicted that, if environmental grit load explains the variation, other large-bodied, terrestrial mammals in the deposits with early hominins should show similar differences in microwear pattern between southern and eastern African samples.

We compared dental microwear textures of alcelaphin and antilopin fossil bovids, along with published data for hominins[16-20] found at the same sites. The bovid data were originally presented in Scott[21] and Steininger[22] (see Appendix 1 in the supplementary material). These tribes were selected because they are common at Plio-Pleistocene fossil sites in both eastern and southern Africa, and because extant representatives have very different dietary patterns.[23,24] Extant alcelaphins are predominantly grazers, although some consume browse when grass is scare. Antilopins, in contrast, include the whole gamut from obligate grazers to obligate browsers; and these differences are clearly reflected in dental microwear texture patterns (Supplementary table 1).

Original specimens were cleaned with alcohol-soaked cotton swabs, and microwear impressions were taken on first or second molar teeth using President's Jet regular body polyvinyl-siloxane dental impression material (Coltène-Whaledent Inc, Cuyahoga Falls, OH, USA). Replicas were poured using Epotek 301 high-resolution epoxy and hardener (Epoxy Technologies, Billerica, MA, USA) and examined for post-mortem damage. Those specimens preserving ante-mortem microwear were scanned using the Sensofar plμ standard white-light confocal profilometer at the University of Arkansas to obtain point clouds representing four adjoining fields. The lateral (x, y) sampling interval was 0.18 μm, vertical (z) resolution was 0.005 μm and field of view for each scan was 138x102 μm.

Observable artifacts, such as dust, were deleted electronically and the point clouds were imported into ToothFrax (Surfract Corp, www.surfract.com) to determine area-scale fractal complexity (*Asfc*) and length-scale anisotropy of relief (*epLsar*). Surfaces with pits of varying sizes tend to have high *Asfc* values, whereas those dominated by aligned scratches have higher *epLsar* values. These variables were chosen because previous studies have shown that browsing bovids have higher average *Asfc* values whereas grazers have higher *epLsar* values.[25-27] In fact, these attributes together effectively parse extant bovids into Gagnon and Chew's[28] fine-scale diet categories[29]: (1) obligate grazers (>90% monocots); (2) variable grazers (60–90% monocots); (3) browser-grazer intermediates (30–70% monocots and dicots, including some fruit); (4) generalists (>20% of each of the three food types); (5) browsers (>70% dicots only, part fruit); and (6) frugivores (>70% fruits).

Median values of the four scans of each specimen were calculated, and the final data set was rank-transformed to mitigate violation of assumptions inherent to parametric study (see Scott et al.[30] for details). A two-factor multivariate analysis of variance (MANOVA) model was used to analyse the data, with location (eastern versus southern Africa) and tribe (Antilopini, Alcelaphini and Hominini) as the factors, and *Asfc* and *epLsar* as the variables.

Results showed a significant interaction between location and tribe, indicating that the pattern of differences between eastern and southern African specimens varied between alcelaphins, antilopins and hominins. There was no significant difference in microwear texture between southern and eastern African alcelaphins (Tables 1 and 2; Figures 1 and 2). On the other hand, there was significant variation by location for both the antilopins and hominins considered. The differences were, however, in opposite directions: antilopins from eastern Africa had higher complexity values on average than those from southern Africa, whereas hominins from southern Africa had higher complexity averages than those from eastern Africa. The hominin pattern holds for both *Australopithecus* and *Paranthropus* samples.[7]

Further, when we superimpose the fossil bovid data on a microwear texture plot for extant species (Figure 3), the ranges of values for extinct alcelaphins and antilopins from southern Africa overlap primarily with variable grazers, whereas that for antilopins from eastern Africa covers much of the extant browser space.

These results indicate that fossil alcelaphins, antilopins and hominins from southern Africa and those from eastern Africa do not show similar differences in microwear textures. Therefore, assuming bovid and hominin foods were subject to the same abrasive environments at sites within these regions, dust or grit alone does not explain the microwear differences observed. Given the mix of grazers and browsers among the fossil bovids, and the combination of C_3 and C_4 isotope signatures of the hominins[31], it seems likely that these taxa overlapped in feeding height and concentration of exogenous abrasives on food. Further, the fact that the distributions of microwear texture values for fossil and extant samples closely approximate one another for both the antilopin and alcelaphin tribes (compare Figure 2 to Supplementary figure 1) further supports the idea that differences between southern and eastern African hominins are not explained by grit or dust load. Finally, while it is possible that differences in masticatory biomechanics, mineralisation and enamel microstructure could complicate interpretations of differences in patterns seen between hominins and bovids, it is unlikely that these explain the differences between the antilopins and alcelaphins given that microwear differences so strongly mirror diet differences in extant species of these tribes (Supplementary figure 1).

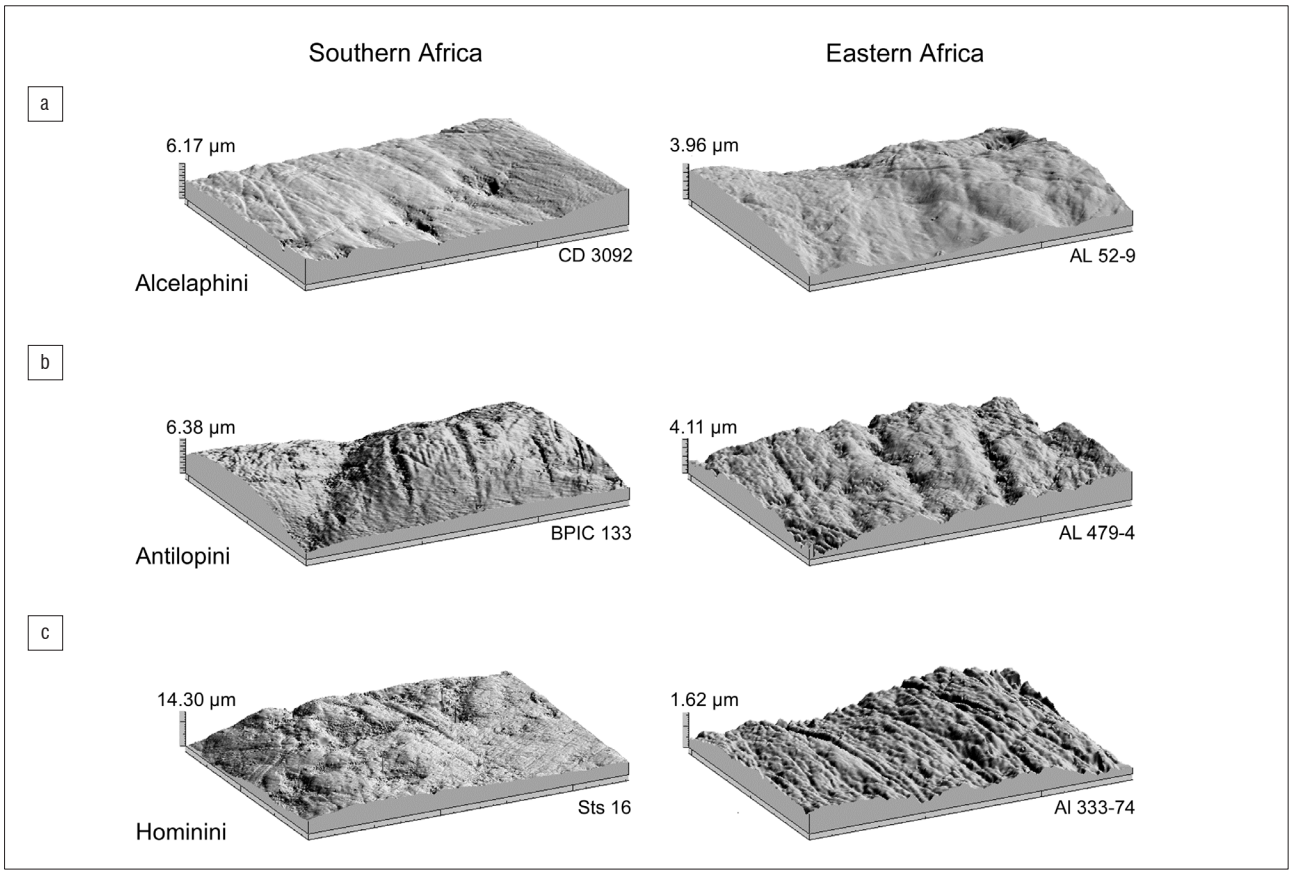


Figure 1: Sample photomicrowear simulations for specimens from southern Africa (left) and eastern Africa (right) representing (a) Alcelaphini, (b) Antilopini and (c) Hominini. Each surface represents an area 138x102 μm.

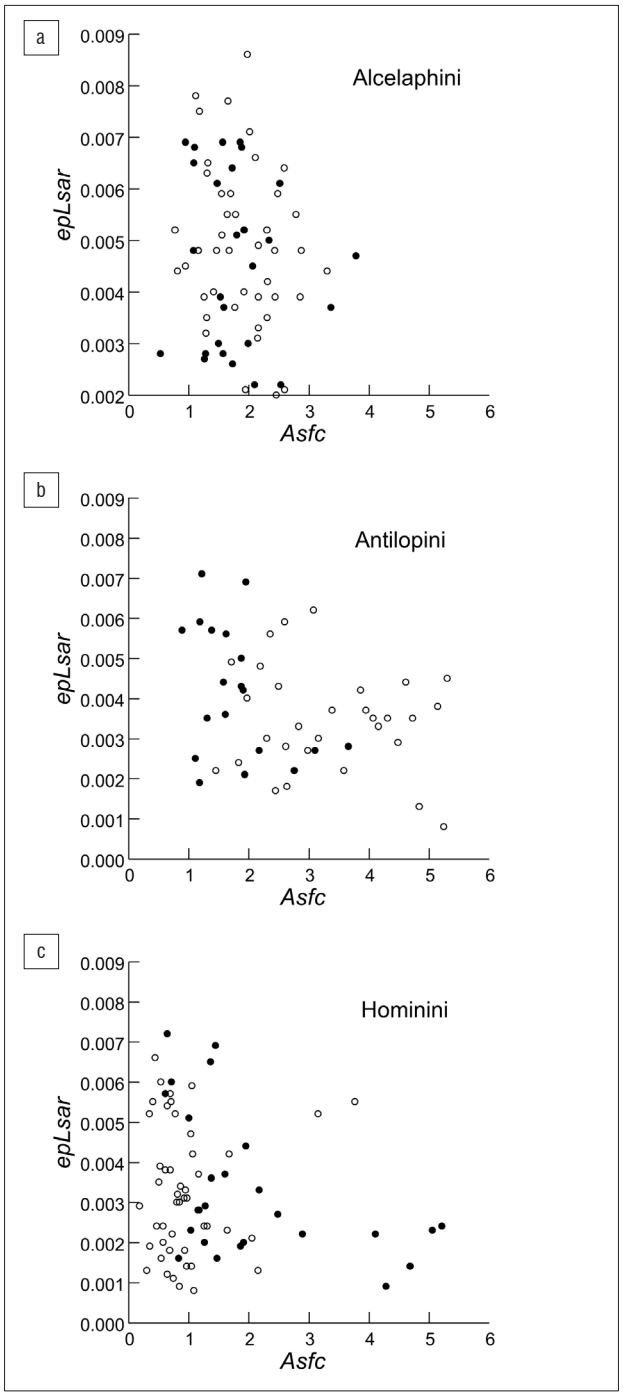


Figure 2: Anisotropy (*epLsar*) versus complexity (*Asfc*) plots for the fossil (a) Alcelaphini, (b) Antilopini and (c) Hominini. Eastern African specimens are clear and southern African ones are solid. Original data for all individual specimens are presented in Appendix 1 in the supplementary material.

Table 1: Mean (s.d.) for microwear attributes by tribe and location

	Eastern			Southern		
	Asfc	*epLsar*	*n*	*Asfc*	*epLsar*	*n*
Alcelaphini	1.873 (0.608)	0.005 (0.002)	44	1.788 (0.701)	0.005 (0.002)	27
Antilopini	3.349 (1.144)	0.003 (0.001)	30	1.772 (0.719)	0.004 (0.002)	20
Hominini	0.964 (0.676)	0.003 (0.002)	47	2.067 (1.417)	0.003 (0.002)	26

Table 2: Summary of results of the multivariate analysis of variance (MANOVA†)

	Value	*F*	d.f.	*p*
Interaction	0.938	3.062	4, 37	0.017
Antilopini	0.979	0.740	2, 68	0.481
Alcelaphini	0.565	18.062	2, 47	0.000
Hominini	0.712	14.171	2, 70	0.000

†*A two-factor MANOVA model was used to analyse the data, with location (eastern versus southern Africa) and tribe (Antilopini, Alcelaphini and Hominini) as the factors, and* Asfc *and* epLsar *as the variables.*

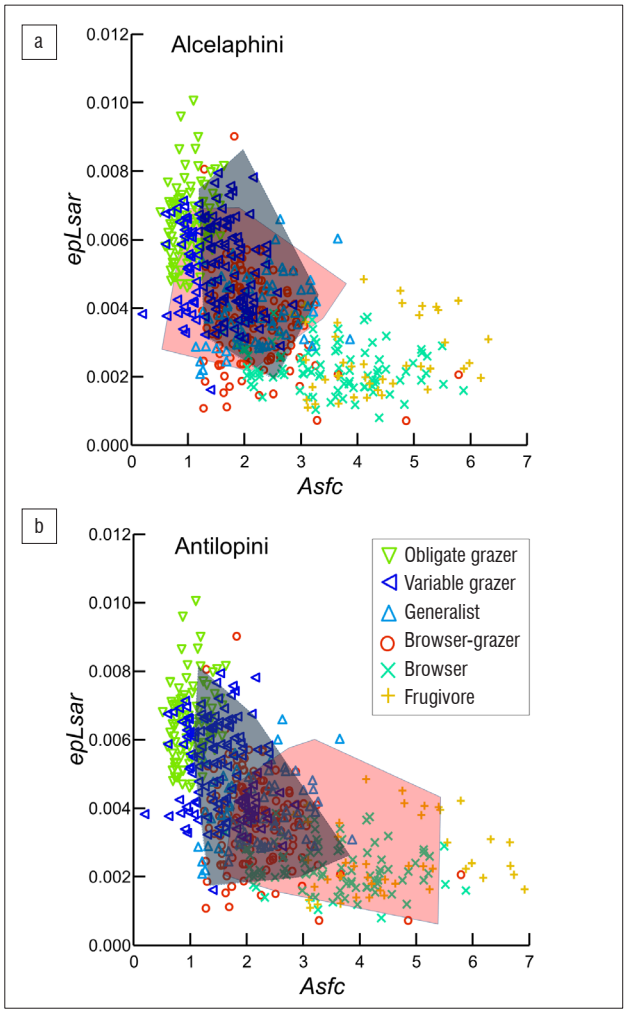


Figure 3: Anisotropy (*epLsar*) versus complexity (*Asfc*) plots for the fossil (a) Alcelaphini and (b) Antilopini compared with baseline specimens, with diets as indicated in the legend (extant bovid data from Scott[29]). The space represented by eastern African fossil specimens is depicted by the red polygon, whereas that represented by the southern African fossil specimens is depicted by the blue polygon. The polygons were constructed by connecting the farthest separated data points on both axes.

In sum, while grit undoubtedly impacts tooth wear[9,13,14], the lack of a consistent location signal among the tribes suggests that differences in microwear between eastern and southern African hominins are likely not a result of abrasive load alone. Diet remains the most plausible explanation for the variation in dental microwear among species. While Sanson et al.[32] and Lucas et al.[9] argued that exogenous grit is

the operative wear agent for teeth because endogenous silicates within plant foods (phytoliths) are softer than enamel, it is clear that there is more to tooth wear than abrasive hardness. Because hydroxyapatite crystallites are attached to one another by a thin layer of protein 'glue', tissue removal requires only that contact pressure be sufficient to break the bonds holding enamel together. Indeed, tissue removal is achieved with particles much softer than enamel.[33] In this light, it makes sense that primates known to consume phytolith-rich foods tend to have thicker tooth enamel[34], that tell-tale siliceous plant opals have been found embedded in tooth enamel at the ends of microwear scratches[35], and that experimental studies show cereals with different phytolith loads leave different microwear patterns[36].

Acknowledgements

The hominin data were collected originally in collaboration with Robert Scott, Fred Grine, Mark Teaford, Kristin Krueger and Alejandro Pérez-Pérez. Funding for data collection came from the US National Science Foundation (P.S.U. and J.R.S.) and the South African National Research Foundation, DST-NRF Centre of Excellence in Palaeosciences and Palaeontological Scientific Trust (C.M.S.). We thank three anonymous reviewers and Francis Thackeray for their helpful comments on an earlier version of this manuscript, and the curators at the various museums who allowed us to study specimens in their care.

Authors' contributions

P.S.U., J.R.S. and C.M.S. gathered the data, analysed the results and wrote the manuscript. P.S.U. was the project leader.

References

1. Purnell MA. Microwear on conodont elements and macrophagy in the first vertebrates. Nature. 1994;374:798–800. http://dx.doi.org/10.1038/374798a0
2. Grine FE. Trophic differences between 'gracile' and 'robust' australopithecines: A scanning electron microscope analysis of occlusal events. S Afr J Sci. 1981;77:203–230.
3. Purnell M, Seehausen O, Galis F. Quantitative three-dimensional microtextural analyses of tooth wear as a tool for dietary discrimination in fishes. J R Soc Interface. 2012;9:2225–2233. http://dx.doi.org/10.1098/rsif.2012.0140
4. Krueger KL. Reconstructing diet and behavior in bioarchaeological groups using incisor microwear texture analysis. J Archaeol Sci Reports. 2015;1:29–37. http://dx.doi.org/10.1016/j.jasrep.2014.10.002
5. Hua L-C, Brandt ET, Meullenet J-F, Zhou Z-R, Ungar PS. Technical note: An in vitro study of dental microwear formation using the *BITE Master II* chewing machine. Am J Phys Anthropol. 2015;158(4):769–775. http://dx.doi.org/10.1002/ajpa.22823
6. Teaford MF, Walker A. Quantitative differences in dental microwear between primate species with different diets and a comment on the presumed diet of *Sivapithecus*. Am J Phys Anthropol. 1984;64:191–200. http://dx.doi.org/10.1002/ajpa.1330640213
7. Ungar PS, Sponheimer M. The diets of early hominins. Science. 2011;334:190–193. http://dx.doi.org/10.1126/science.1207701
8. Strait DS, Constantino P, Lucas PW, Richmond BG, Spencer MA, Dechow PC, et al. Viewpoints: Diet and dietary adaptations in early hominins: the hard food perspective. Am J Phys Anthropol. 2013;151:339–355. http://dx.doi.org/10.1002/ajpa.22285
9. Lucas PW, Omar R, Al-Fadhalah K, Almusallam AS, Henry AG, Michael S, et al. Mechanisms and causes of wear in tooth enamel: Implications for hominin diets. J R Soc Interface. 2013;10(80), Art. #20120923, 7 pages. http://dx.doi.org/10.1098/rsif.2012.0923
10. Wood B. Palaeontology: Gritting their teeth. Nature. 2013;493:486–487. http://dx.doi.org/10.1038/493486a
11. Ungar PS. Mammalian dental function and wear: A review. Biosurf Biotribol. 2015;1:25–41. http://dx.doi.org/10.1016/j.bsbt.2014.12.001
12. Williams FL. Dietary proclivities of *Paranthropus robustus* from Swartkrans, South Africa. Anthropol Rev. 2015;78:1–19. http://dx.doi.org/10.1515/anre-2015-0001
13. Galbany J, Romero A, Mayo-Alesón M, Itsoma F, Gamarra B, Pérez-Pérez A, et al. Age-related tooth wear differs between forest and savanna primates. PLoS One. 2014;9(4), e94938, 7 pages. http://dx.doi.org/10.1371/journal.pone.0094938
14. Madden R. Hypsodonty in mammals: Evolution, geomorphology, and the role of earth surface processes. Cambridge: Cambridge University Press; 2015.
15. Gomes Rodrigues H, Merceron G, Viriot L. Dental microwear patterns of extant and extinct Muridae (Rodentia, Mammalia): Ecological implications. Naturwissenschaften. 2009;96:537–542. http://dx.doi.org/10.1007/s00114-008-0501-x
16. Scott RS, Ungar PS, Bergstrom TS, Brown CA, Grine FE, Teaford MF, et al. Dental microwear texture analysis shows within-species diet variability in fossil hominins. Nature. 2005;436:693–695. http://dx.doi.org/10.1038/nature03822
17. Ungar PS, Grine FE, Teaford MF. Dental microwear and diet of the Plio-Pleistocene hominin *Paranthropus boisei*. PLoS One. 2008;3(4), e2044, 6 pages. http://dx.doi.org/10.1371/journal.pone.0002044
18. Ungar PS, Scott RS, Grine FE, Teaford MF. Molar microwear textures and the diets of *Australopithecus anamensis* and *Australopithecus afarensis*. Phil Trans R Soc Lond. B. 2010;465:3345–3354. http://dx.doi.org/10.1098/rstb.2010.0033
19. Ungar PS, Krueger KL, Blumenschine RJ, Njau J, Scott RS. Dental microwear texture analysis of hominins recovered by Olduvai Landscape Paleoanthropology Project, 1995-2007. J Hum Evol. 2012;63:429–437. http://dx.doi.org/10.1016/j.jhevol.2011.04.006
20. Ungar PS, Scott RS. Dental evidence for diets of early *Homo*. In: Grine FE, Leakey RE, Fleagle JG, editors. The first humans: Origins of the genus *Homo.* New York: Springer-Verlag; 2009. p. 121–134. http://dx.doi.org/10.1007/978-1-4020-9980-9_11
21. Scott JR. Dental microwear texture analysis of Pliocene bovids from four early hominin fossil sites in eastern Africa: Implications for paleoenvironmental dynamics and human evolution [PhD dissertation]. Fayetteville, AR: University of Arkansas; 2011.
22. Steininger CM. Dietary behaviour of early Pleistocene bovids from Cooper's Cave and Swartkrans, South Africa [PhD thesis]. Johannesburg: University of the Witwatersrand; 2011.
23. Estes R. The behavior guide to African mammals: Including hoofed mammals, carnivores, primates. Berkeley and Los Angeles, CA: University of California Press; 1991.
24. Skinner J, Chimimba C. The mammals of the southern African subregion. Cambridge: Cambridge University Press; 2005. http://dx.doi.org/10.1017/CBO9781107340992.014
25. Ungar PS, Merceron G, Scott RS. Dental microwear texture analysis of Varswater bovids and early Pliocene paleoenvironments of Langebaanweg, Western Cape Province, South Africa. J Mamm Evol. 2007;14:163–181. http://dx.doi.org/10.1007/s10914-007-9050-x
26. Merceron G, Escarguel G, Angibault J-M, Verheyden-Tixier H. Can dental microwear textures record inter-individual dietary variations? PLoS One. 2010;5(3), e9542, 9 pages. http://dx.doi.org/10.1371/journal.pone.0009542
27. Schulz-Kornas E, Calandra I, Kaiser TM. Feeding ecology and chewing mechanics in hoofed mammals: 3D tribology of enamel wear. Wear. 2013;300:169–179. http://dx.doi.org/10.1016/j.wear.2013.01.115
28. Gagnon M, Chew AE. Dietary preferences in extant African Bovidae. J Mammal. 2000;81(2):490–511. http://dx.doi.org/10.1644/1545-1542(2000)081<0490:DPIEAB>2.0.CO;2
29. Scott JR. Dental microwear texture analysis of extant African Bovidae. Mammalia. 2012;76:157–174. http://dx.doi.org/10.1515/mammalia-2011-0083
30. Scott RS, Ungar PS, Bergstrom TS, Brown CA, Childs BE, Teaford MF, et al. Dental microwear texture analysis: Technical considerations. J Hum Evol. 2006;51:339–349. http://dx.doi.org/10.1016/j.jhevol.2006.04.006
31. Sponheimer M, Alemseged MZ, Cerling TE, Grine FE, Kimbel WH, Leakey MG, et al. Isotopic evidence of early hominin diets. Proc Natl Acad Sci USA. 2013;110(26):10513–10518. http://dx.doi.org/10.1073/pnas.1222579110
32. Sanson GD, Kerr SA, Gross KA. Do silica phytoliths really wear mammalian teeth? J Archaeol Sci. 2007;34:526–531. http://dx.doi.org/10.1016/j.jas.2006.06.009

33. Xia J, Zheng J, Huang D, Tian ZR, Chen L, Zhou Z, et al. New model to explain tooth wear with implications for microwear formation and diet reconstruction. Proc Natl Acad Sci USA. 2015;112:10669–10672. http://dx.doi.org/10.1073/pnas.1509491112

34. Rabenold D, Pearson OM. Abrasive, silica phytoliths and the evolution of thick molar enamel in primates, with implications for the diet of *Paranthropus boisei*. PLoS One. 2011;6(12), e28379, 11 pages. http://dx.doi.org/10.1371/journal.pone.0028379

35. Lalueza Fox C, Pérez-Pérez A. Dietary information through the examination of plant phytoliths on the enamel surface of human dentition. J Archaeol Sci. 1994;21:29–34. http://dx.doi.org/10.1006/jasc.1994.1005

36. Gügel IL, Grupe G, Kunzelmann KH. Simulation of dental microwear: Characteristic traces by opal phytoliths give clues to ancient human dietary behavior. Am J Phys Anthropol. 2001;114:124–138. http://dx.doi.org/10.1002/1096-8644(200102)114:2<124::AID-AJPA1012>3.0.CO;2-S

Note: This article includes supplementary material