SOUTH AFRICAN Journal of Science

volume 110
number 3/4

Geothermal energy potential in South Africa

Elephants in Africa: Biodiversity thieves?

Lower emissions from shale gas for electricity generation

Higher education's contribution to the South African economy

MARCH/APRIL 2014

ISSN: 0038-2353

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

MARCH/APRIL

ISSN: 0038-2353 (Print) | ISSN: 1996-7489 (Online)

SOUTH AFRICAN Journal of Science

volume 110
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 4300 459
E: admin@sunbloem.co.za

Correspondence and enquiries
sajs@assaf.org.za

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

Subscriptions
Subscription information can be found at www.sajs.co.za

Research Article

Research Letter

Changes in the offing

It comprises 411 pages, has a Ministerial Forward, a Chairperson's Introduction and an Executive Summary of well over 14 000 words. Add in some nifty formulae and graphs and you have the *Report of the Ministerial Committee for the Review of the Funding of Universities.* To capture and summarise the detailed content of so substantial a report in a thousand words or so would be an impossible task. But there are four aspects of the *Report* that might be of interest to readers of this Journal: the review team's recommendations will be modelled to determine their effects on South African universities, emphasis is placed on steering the higher education system towards differentiation, there are proposed changes to the rewards for publication and a proposed shift in focus in the allocation of Research Development Grants.

In order to test the recommendations made in the *Report*, a technical team and parallel reference group have been established to determine what those recommendations will mean for each of the 25 universities in the system. Due to be completed by the end of 2014, the outcome will form the basis for a draft funding policy and framework which, after consultation with the sector, will become the basis for future funding. It is intended that the new policy and framework will then be phased in over a number of years – although a so-called 'disadvantage factor' will be introduced to the existing funding framework and implemented immediately, using new funds, to address the weak financial circumstances of some historically (and currently) disadvantaged institutions.

The emphasis on creating and maintaining a differentiated higher education system develops the idea set out in the White Paper on Post School Education and Training released in January this year. In the White Paper's section on universities, the point is made that

> [t]*here is broad agreement that South Africa needs a diverse university sector which is purposefully differentiated in order to meet a range of social, economic and educational requirements. … Since the establishment of the* [Department of Higher Education and Training]*, it has been recognised that the principle of differentiation must apply beyond the universities to the entire post-school system.*

In his Foreword to the *Report*, Minister Blade Nzimande takes the point further:

> *A number of our universities are world-class academic institutions at the cutting edge of research in various spheres, while others may be better situated to make teaching their primary purpose. This brings the need for a differentiated university sector to the fore. Both teaching and research are critical for the development of highly skilled academics, workers and researchers, and it is important that both these activities are adequately funded. It is government's vision that all universities should at least develop research niche areas and that all universities will participate in research and innovation, albeit to various extents.*

The recommendations set out in the *Report* are quite specific:

> *The most important principle is that the country needs the entire spectrum of institutions for socio-economic development. The higher education sector should comprise a continuum of institutions, with the purpose of providing a varied student population with a range of access routes. … A variety of institutions are* [sic] *therefore required, to ensure that the sector serves the varied needs of students as well as the national interests. The mix and level of programmes of any institution should not be cast in stone; institutions must identify and enhance their strengths.*

It is unclear as to how well this will be received by some vice-chancellors.

The position taken by the Review Committee on financial rewards for publications starts on a positive note, pointing out that the funds available for research productivity are limited and that '[i]t is therefore essential that more funding be allocated by Treasury to ensure that the success in increasing research outputs of universities be sustained and further incentivised'. There are, however, recommended changes regarding how research publications should be assessed, with an emphasis on rewarding excellence and quality rather than quantity. To achieve this aim, accredited journals would be ranked and assigned to three categories based on their impact factors published by, for example, the Web of Science. Publication rewards would then be based on the category of journal in which research is published. The proposal does, however, pose some problems: no account is taken of the fact that 'high' impact factors vary widely among disciplines, nor that many South African based journals do not appear amongst the published citation indexes. It is also recommended that papers which are the products of collaborative research, whether nationally or regionally, should receive higher rewards than papers that do not involve collaboration.

In order to support emerging scholars, it is suggested that two new journals might be established (one in the humanities and social sciences and a second in the natural sciences) with the specific aim of carrying first-time papers published by postgraduate students – and it is recommended that these journals would best be administered by the Academy of Science of South Africa. Although it is unclear what is meant by 'research outputs from the performing, creative and visual arts,' the *Report* recommends that these should also be assessed and, where appropriate, suitably funded.

Finally (for this Leader), there is the matter of Research Development Grants. The *Report* indicates that 'there is evidence' to suggest that the current, ring-fenced Research Development Grants are not being used effectively and that the focus should shift to universities and universities of technology that have limited research capacities, with funding aimed at developing a critical number of research niche areas around centres of excellence. Such a shift would, the *Report* argues, require a careful analysis of the research strengths of targeted institutions with an emphasis on the status of research development strategies and plans. In general, however, the view is expressed that the goal of building research capacity for the entire system should be re-affirmed and that special attention should be given to support for emerging scholars – with the implication that such support should be a factor included in general funding as it will be defined in the new policy and framework.

HOW TO CITE: Butler-Adam J. Changes in the offing. S Afr J Sci. 2014;110(3/4), Art. #a0061, 2 pages. http://dx.doi.org/10.1590/sajs.2014/a0061

Towards the end of his Introduction to the *Report*, the Chairperson of the Ministerial Committee, Mr Cyril Ramaphosa, observes that

> *South Africa's funding of higher education, even though significant, does not compare favourably to other countries. In 2011, the state budget for universities as a percentage of gross domestic product (GDP) was 0.75% compared to 0.78% for Africa as a whole, and 1.21% for the Organisation for Economic Co-operation and Development (OECD) countries. Given the important role of higher education in the production of skills, research and innovation, in the mitigation of socio-economic inequalities, and in the realisation of the state's development agenda, the level of funding needs to improve. However, for additional funding to have any meaningful impact, it is necessary to address the inefficiencies in the system.*

Let us hope that the final version of the funding policy and framework will help to address the inefficiencies and allow for that much needed additional injection of financial support.

Is the Equity Index a good tool to gauge demographic transformation at South African universities?

AUTHOR:
Alan G. Morris[1]

AFFILIATION:
[1]Department of Human Biology, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Alan Morris

EMAIL:
alan.morris@uct.ac.za

POSTAL ADDRESS:
Department of Human Biology, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa

KEYWORDS:
Equity Index; South Africa; university, demographic transformation

HOW TO CITE:
Morris A. Is the Equity Index a good tool to gauge demographic transformation at South African universities? S Afr J Sci. 2014;110(3/4), Art. #a0056, 2 pages. http://dx.doi.org/10.1590/sajs.2014/a0056

The article by Govinder et al. (S Afr J Sci. 2013;109(11/12)) on assessing the demographic transition in South African universities is extremely interesting and is certainly something that university administrators around the country will find valuable. During my own tenure at the University of Cape Town (UCT), from 1981 to the present, I have seen the student demography change dramatically. The MBChB intake that I saw in my first year teaching at UCT was overwhelmingly white and male, whereas the current intakes are predominantly female, and white students represent less than 40% of the students enrolled. The precise figures for all undergraduate students at UCT in 2011 were 36% black, 20% coloured, 10% Indian and 31% white.[1] The transformation took place in two waves: the first after 1983 when the government dropped its racial controls on admissions and the second from 1990 when UCT began to actively recruit and select African students.[2] During this period, UCT has been criticised for using racial identity as a factor in the selection of students and the medical school in particular has had its admissions policy challenged in the courts and at the Human Rights Commission.[3] The outcome of the legal cases has been that the University's right to use race as a selection factor has been upheld because changing the demography of the student body is an important aspect of righting the wrongs of the past. Despite the need for demographic transformation in higher education and the obvious value of the Equity Index created by Govinder et al. in assessing transformation, several issues in their analysis suggest that more thought is required before the index is applied as an intervention tool.

The data presented by Govinder et al. indicate that not all universities in South Africa have completed the transformation process. Certainly this is true based on national demographic statistics, but national demographic ratios might not be the best standard to work with. The authors baldly state that the national data are the preferred benchmark because 'all South African universities are classified as national assets'. Such a rigid approach has already been challenged in the courts and found to be wanting. Judge Rabkin-Naicker ordered the Department of Correctional Services to take immediate steps to ensure that both national and regional demographics were taken into account in respect of members of designated groups when setting equity targets at all occupational levels of its workforce.[4] Although the legal issue is important, there is an even more compelling reason to consider regional rather than national demography. Students will preferentially choose a university that is closer to home if that is possible. Living at home and commuting to school on a daily basis is clearly a financial advantage, but even when students are in residence, there is a preference to be closer to home if possible so that visits to family are easier and can be made more frequently. This factor is not considered in the analysis and it does help to explain why all four Western Cape institutions fall at the 'worse' end of the equity index. I would be fascinated to see how the equity indices would appear if provincial demographic data were used instead of national data.

The data from the Equity Efficiency Index do tell us that the match to national demography is poorer in graduating classes than for admissions. Although no breakdown is given in the article, my own experience at UCT is that African students struggle and that even with substantial resources being spent to help these students cope, their failure rate remains higher than students in other demographic categories. Data from UCT indicate that of the 2007 admissions cohort, only 48% of black, 67% of coloured, 68% of Indian, but 81% of white students graduated within 5 years. The situation had actually worsened for black students, as nearly 60% had graduated from the 2003 cohort.[1] Meetings have been convened at all levels from departmental to University Senate to look at the cause of this problem and many factors have been identified. The most important factor by far is the language barrier for students who speak English as a second language. This disproportionately affects black African students. Add to this the fact that minimum academic scores for admission are lower for African students precisely because UCT wants to admit more African students. Other factors that impact on success rate are culture shock, lack of educational triggers in the pre-university home environment, and inadequate primary and secondary preparation, none of which can be easily fixed at university. I can only speak for my UCT Faculty of Health Sciences experience, but we have had an extensive intervention programme in place for years and my own Department of Human Biology has three academic staff members who have been hired specifically to engage with the intervention programme students.

In the light of this considerable effort to help weaker students, I am particularly concerned that Govinder et al. think that a positive Equity Efficiency Index is something desirable. The target Equity Efficiency Index must be zero, and any deviation, either positive or negative, suggests that we are failing to provide proper support for students somewhere along the line. Intervention programmes must not be applied on the basis of race, but only on the need for academic help, so the worst possible outcome would be a positive index because that would mean that resources are being applied on the basis of race rather than academic need.

Lastly is the hoary issue of the relationship between transformation and research output. There is no question that the demographic transformation of academic staff has been much slower than the transformation of the student body. Govinder et al. do not emphasise this disparity, but a simple comparison of the average Equity Index for student admissions (32.87) with that for academic staff (63.3) shows how far apart they are. Despite this disparity, a simple correlation of these two data sets produces a significant r-value of 0.562, showing that, at least in proportional terms, the demography of the academic staff is broadly similar to the student intake in each institution. This finding is important to note as the demographics of the students *have* transformed at all universities and therefore so have academic staff even if they do not match the national demographic ratios.

Of greater importance is the relationship of research output (a form of quality assessment) with staff equity. Govinder et al. tell us that there 'is no direct linear correlation between EI ranking and research productivity', but this is not true. The *r* correlation coefficient for their two data sets is a statistically significant 0.588. When the per capita research output is considered, the *r*-value rises to 0.745. It is always unwise to assign cause and effect to significant correlations, but the data indicate that the two variables are connected and that poor equity scores pair with higher research output and vice versa. The universities fall into three groups (with the University of Fort Hare as an outlier) and it is indeed disturbing to see how five 'previously advantaged' schools in the graph (Wits, Pretoria, Rhodes, UCT and Stellenbosch) separate out. These schools are not just separate from other South African schools, but actually fall well within the spectrum of top-class universities on a world scale.[5,6]

One almost certainly false interpretation of this result would be that academic staff transformation damages research quality. A more realistic interpretation is that different universities have different priorities in hiring and promotion of academic staff. The cluster of 'previously advantaged' universities base their hiring and promotion standards on high levels of research production, while the cluster of technical universities in the opposite quadrant of the graph have traditionally focused on teaching in terms of hiring and promotion. This means that the different universities are drawing academic staff from different pools of applicants with differing demographic structures (especially in relation to overseas applicants). The pool of skilled scholars from previously disadvantaged demographic categories in South Africa is growing, but it is not yet at the point at which all applicants for high-level research posts fit the national demographic profile.

My worry is that Govinder et al. do not recognise these issues in their discussion and promotion of the Equity Index. Transformation is only seen in the light of racial and gender equity and it is implied that universities which do not meet the national demographic ratios for staff or students need to find ways to correct this. It is a given that all South African universities still require demographic transformation for both staff and students to a greater or lesser extent, but is the solution to use only equity issues when planning for the future? Govinder et al. ask whether the slow pace of demographic change in the universities is because of active resistance to transformation by the 'privileged' or passive resistance through conservatism. They ask if it is time to reconsider self-regulation and autonomy, and by implication suggest the possible implementation of demographic quotas in admission and hiring. Certainly in my experience there has been no active resistance to transformation. Change has already been marked, just not in the rigid direction that Govinder et al. demand. I wonder how they will respond to the new proposals for admission to UCT from 2015 in which race is removed as the central factor in admissions.[7] The proposals will give higher admission credits to students from rural contexts, and those with families with first time exposure to university education and lower socio-economic status. These factors may or may not overlap with racial categories, but their inclusion is an attempt to broaden the transformation at UCT by taking the focus away from the 'privileged' few by concentrating on social class rather than on race.

On a last note, I have sat on the UCT Faculty of Health Sciences Readmission Review Committee for over a decade. This committee is the 'last chance' committee that considers appeals by students who have been excluded from the Faculty on academic grounds. It is by far the most difficult and emotionally taxing committee on which I have served and each and every student that is reviewed is considered purely on the basis of the committee members' belief in whether or not the student will ultimately succeed if readmitted. Race or gender is never an issue.

References

1. Teaching and Learning Report to UCT Senate; University of Cape Town; 12 September 2012. Available from: *www.uct.ac.za/usr/ipd/IIU/intreports/tlreport/T&l2011NAR.pdf*

2. Morris AG. The Human Biology course for MBChB1: A retrospective, 1982 to 2001. 2004 [unpublished].

3. South African Human Rights Commission Case WC/2008/0411 [case on the Internet]. 05 May 2011 [cited 2014 Mar 12]. Available from: *www.sahrc.org.za/home/21/files/Finding%20of%20Kemp.pdf*

4. Labour Court of South Africa, Judgment Case no: C368/2012 [case on the Internet]. 18 Dec 2013 [cited 2014 Mar 12]. Available from: http://www.saflii.org.za/za/cases/ZALCCT/2013/38.html

5. QS World University rankings [document on the Internet]. c2013 [cited 2014 Mar 12]. Available from: http://www.topuniversities.com/university-rankings/world- university-rankings/2013

6. Times Higher Education World University Rankings [document on the Internet]. c2014 [cited 2014 Mar 12]. Available from: http://www.timeshighereducation.co.uk/world-university-rankings/2013-14/world-ranking

7. Report of the Commission into Student Admissions. Special Senate Agenda, University of Cape Town; 15 March 2013 [unpublished report].

Value measurement theory and league tables

AUTHOR:
Theodor J. Stewart[1]

AFFILIATION:
[1]Emeritus Professor, Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Theodor Stewart

EMAIL:
theodor.stewart@uct.ac.za

POSTAL ADDRESS:
Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, Cape Town, South Africa

KEYWORDS:
multicriteria decision analysis; university rankings; scoring systems

HOW TO CITE:
Stewart TJ. Value measurement theory and league tables. S Afr J Sci. 2014;110(3/4), Art. #a0057, 2 pages. http://dx.doi.org/10.1590/sajs.2014/a0057

This communication was stimulated by the editorial in the Sep/Oct 2013 issue of the *South African Journal of Science* concerning university rankings, which rang a strong sympathetic chord with me. I have for a long time felt a professional responsibility as a decision scientist or decision analyst (specialising in multicriteria decision analysis) to respond to the irresponsible use of scoring systems, league tables, etc., which so many people accept uncritically.

From a decision analytic point of view, most ranking systems are based on a simple multicriteria value function, typically additive in structure. In other words, for each alternative to be ranked (largely universities in this discussion), a score is calculated of the form $V_a = \sum_{i=1}^{m} W_i V_{ai}$, where W_i is the importance weight associated with criterion *i* (out of *m*), and V_{ai} is the evaluation (or partial score) achieved by alternative *a* with respect to criterion *i*.

There exists a substantial theory on the validity and properties of such value functions (which are reviewed for example in Belton and Stewart[1]), but the developers of ranking systems show little evidence of awareness of this theory. In particular, there are two assumptions which are critical to the validity of additive value functions:

- The criteria on which they are based must be *preferentially independent*. This is a frequently misunderstood condition and has nothing to do with statistical independence; it is the ability to express trade-offs between two criteria, on the assumption that performances of all other criteria are the same, independent of the levels of these constant performances. It is not evident that preferential independence assumptions have received much, if any, attention by developers of university ranking schemes.
- The scores on each criterion must satisfy the properties of an *interval scale of preference*, so that the differences in scores have the same importance irrespective of the baseline level. Consider for example two alternatives (universities), say *a* and *b*. These would achieve equal ranks if and only if $\sum_{i=1}^{m} W_i [V_{ai} - V_{bi}] = 0$ which depends entirely on the differences $V_{ai} - V_{bi}$ and not on their absolute values. Natural measurements do not naturally satisfy this interval preference scale property – a change in publication count, for example, from 100 to 150 pa cannot in general be associated with the same increase in preference as a change from 1000 to 1050. The definition and mode of assessment of the partial scores V_{ai} can thus be critical to the validity of an additive model (which is discussed below).

It is self-evident (I hope) that no ranking system is an objective measure of performance. It includes a variety of value judgements that should be assessed separately by any users, without uncritically accepting the values used by publishers of rankings. This subjectivity is perhaps quite widely recognised (if not acted upon) in regard to the importance weightings W_i. But sensitivity to the definition of the partial scores V_{ai} is not widely recognised, and yet has been shown in a number of studies (e.g. Stewart[2,3]) to have potentially even larger impacts on output rankings generated by the additive model than that of variations in the importance weights.

Chapter 5 of Belton and Stewart[1] details the processes that need to be undertaken in order to establish defensible assessments of the partial scores. These processes are quite demanding, and developers of league tables resort to simpler approaches. Two common approaches are to select some easily available attribute related to the criterion of interest (e.g. numbers of publications as a surrogate for research output) and to set V_{ai} equal to a linearly scaled value of the attribute or to simply rank order the universities according to the criterion, and then equate the partial scores to quantile values for an assumed population distribution (e.g. the so-called *Z*-scores assuming a Gaussian distribution). There are no credible grounds for assuming that scores obtained in either of the above manners correspond to the required interval scale for preferences, so validity of any derived rankings must be questioned.

As illustration of the last claim, the use of *Z*-scores (quantiles of a Gaussian distribution) for defining partial scores is examined. This approach would be valid for a large number of alternatives if and only if the distribution of values to decision-makers was normal for the population being assessed. No argument from a central limit theorem makes sense here – when looking at the specific population being ranked. On prima facie grounds I find the following implications of *Z*-scores from a normal distribution to be highly implausible as rules to be generally applied in all cases:

- The gain in moving from the alternative appearing as the 25th quantile in the population to the median alternative is deemed to be equivalent to the gain in moving from the median to the 75th quantile
- The gain in moving from the alternative appearing as the 90th percentile to the 95th is deemed to be less than half the gain in moving from the 95th to the 99th

In my experience of fitting value functions by the approaches described in Chapter 5 of Belton and Stewart[1], non-symmetrical value functions are typical (invalidating the first implied property), while the top few alternatives are often not strongly distinguished (invalidating the second). The onus must be on the developers of the ranking system to provide valid evidence for the preference gaps being assumed.

About all that can be inferred with complete justification from the data shown in most league tables is that if one alternative (university) dominates another, in the sense of being better on all criteria, then it is the better of the two (although even here we must assume that no important criteria have been omitted). But, in general, such dominance properties will not provide particularly complete rankings – comparisons between most pairs of alternatives would be indeterminate. My conclusion is that extreme caution and scepticism needs to be applied to league tables, such as university rankings, before they are used for any significant decision-making. This caveat applies equally well to students (and their families) choosing universities, and to university administrators either in boasting about improved league positions or in allocating resources to improving league position.

References

1. Belton V, Stewart TJ. Multiple criteria decision analysis: An integrated approach. Boston, MA: Kluwer Academic Publishers; 2002. http://dx.doi.org/10.1007/978-1-4615-1495-4

2. Stewart TJ. Use of piecewise linear value functions in interactive multicriteria decision support: A Monte Carlo study. Manage Sci. 1993;39:1369–1381. http://dx.doi.org/10.1287/mnsc.39.11.1369

3. Stewart TJ. Robustness of additive value function methods in MCDM. J Multi-Crit Decis Anal. 1996;5:301–309. http://dx.doi.org/10.1002/(SICI)1099-1360(199612)5:4<301::AID-MCDA120>3.0.CO;2-Q

Elephants in Africa: Big, grey biodiversity thieves?

AUTHORS:
Timothy R. Kuiper[1]
Daniel M. Parker[1]

AFFILIATION:
[1]Wildlife and Reserve Management Research Group, Department of Zoology and Entomology, Rhodes University, Grahamstown, South Africa

CORRESPONDENCE TO:
Timothy Kuiper

EMAIL:
timothykuiper@gmail.com

POSTAL ADDRESS:
Wildlife and Reserve Management Research Group, Department of Zoology and Entomology, Rhodes University, PO Box 94, Grahamstown 6140, South Africa

KEYWORDS:
conservation; ecosystem change; habitat heterogeneity; large herbivores; Africa

HOW TO CITE:
Kuiper TR, Parker DM. Elephants in Africa: Big, grey biodiversity thieves? S Afr J Sci. 2014;110(3/4), Art. #a0058, 3 pages. http://dx.doi.org/10.1590/sajs.2014/a0058

At the outset

The conservation of biological diversity is one of the greatest challenges facing humanity today.[1,2] Addressing this challenge must inevitably involve the maintenance of the composition, structure and functioning of multi-species ecosystems.[2-4] While preventing the loss of particular species is an obvious strategy, a second goal might be to manage for local species diversity and the heterogeneity of habitats.[5] A further objective may be the restoration of damaged ecosystems.[6] Large herbivore assemblages form an important component of many diverse ecosystems and are of distinct ecological, and hence conservation, value.[7-9] On the other hand, herbivores managed at unnaturally high densities may drive detrimental changes in the structural heterogeneity of habitats.[10-12] Thus, decision-makers are faced with the challenge of balancing the needs of populations of large herbivores with the preservation of vegetation and ecosystem diversity.[9] In light of these considerations, conclusions as to whether large herbivores are 'good' or 'bad' for biological diversity are both contentious and elusive. Nonetheless, we seek to explore this question with particular reference to African savannah elephants (*Loxodonta africana*). We discuss whether or not the elephant should be considered a biodiversity thief – a species that upsets the natural diversity of life in the habitat in which it lives.

The African elephant has been driven to local extinction over much of its former range, with hunting for meat and ivory and the conversion of elephant habitat to agriculture constituting the major drivers.[5] Given the historical decline in elephant numbers, the conservation of this species appears to be a sensible objective.[9,13] However, the protection afforded to elephants in confined (usually fenced) reserves across Africa has distinct implications for the management of local species diversity and the heterogeneity of habitats.[14,15] Elephants have been shown to have clear impacts on the structure of vegetation, particularly in woodland habitats[12,16,17], and these changes may have knock-on effects for sympatric species[10,18,19]. The effects of elephants on biological diversity in protected areas are of particular concern in light of how expansion in human populations, and the land-use change that follows, places increasing pressure on reserves to preserve biological diversity.[9,10]

Change within ecosystems is natural and inevitable and large herbivores have been living in an ever-changing environment for millions of years.[20] Changes brought on by elephants, therefore, should not automatically be categorised as undesirable, and short-term changes may be part of an overarching trend of long-term stability.[8,21] Furthermore, current elephant impacts can only be adequately assessed in the light of historical (on the scale of centuries) benchmarks, for which accurate information is notoriously scant.[10,21] The current impacts of elephant populations on vegetation dynamics in some regions may in fact represent a shift towards a more natural historical state.[8,21] Finally, much of the detrimental impacts of elephants on habitat structure and diversity, particularly at smaller scales, has resulted from human interventions (such as the erection of fences and the establishment of artificial water points) that have unnaturally confined elephants to localised areas, disrupted seasonal movements and disturbed the natural dynamics of population processes.[5,13,22]

The context of historical change

The demonstrated changes in woodland and riparian forest habitats in northern Botswana, supposedly driven by an expanding elephant population, have raised a considerable amount of concern amongst ecologists in recent years.[17,23-25] Seminal research in northern Botswana's Chobe river ecosystem, while clearly validating elephant-mediated vegetation changes, nevertheless questions the interpretation of these changes as undesirable.[8] The authors argue that the current state of the ecosystem could perhaps be likened to the situation before the historical decline in elephant and other herbivore numbers following excessive hunting for ivory (1800s) and the rinderpest pandemic (c. 1897).[8] If current impacts, however conspicuous, are simply shifting the system back to its former state, then there is no need for concern. Moreover, Owen-Smith et al.[26] stress that in Kruger National Park (KNP), the historical state of vegetation assumed its form in the absence of elephants and other large herbivores (following hunting) and hence changes in vegetation with the recovery of large herbivores are inevitable and not necessarily detrimental. Furthermore, it is difficult to disentangle the relative contribution of fire and elephants to contemporary vegetation change in KNP.[26] Records of past elephant numbers and movement patterns, together with data on historical vegetation structure and composition are essential benchmarks against which contemporary change can be measured.[8] Unfortunately, retracing the history of natural systems is often speculative and records from the 18th and 19th centuries lack any real detail: they consist of the writings of early hunters, archaeological faunal remains and rock paintings.[8,10,27]

The role of elephants in habitat change

Vegetation structure and composition forms an integral component of ecological diversity and, *sensu stricto*, negative effects of elephants on vegetation may therefore be seen as bad for biological diversity.[3,9] Of all habitat types, the impact of elephants on woodland is most obvious and forms the basis of much emotive debate.[12,13,16,17]

In KNP (South Africa), Asner and Levick[12] showed that elephants were the primary agents of treefall and that rates of treefall averaged six times greater in areas accessible to elephants than in exclosures. In Tanzania's Ruaha National Park, elephants were implicated in the death of 37% and 67% of measured *Acacia albidia* and *Commiphora ugogensis* trees, respectively.[16] In the Miombo woodlands of Zimbabwe, the density and cover of trees were significantly lower in elephant-impacted woodlands than in adjacent intact woodland.[10] In Amboseli National Park in Kenya, a long-term series of exclusion experiments showed that elephants alone are preventing the regeneration of woodlands.[11] A similar study in northern Botswana showed that elephants play an important role in reducing the extent of tree cover.[17] In South Africa's Tembe Elephant Park, elephants had a clear influence on vegetation at the species level but had no apparent impact at the woodland community level.[15]

Guldemond and van Aarde[28] carried out a meta-analysis of 238 studies conducted over 45 years, all of which investigated the impacts of elephants on vegetation in some way. The majority of the most commonly cited and influential studies (15 of 20) recorded negative impacts of elephants on vegetation while 53% of the remaining 218 studies also recorded negative impacts.[28] There may have been introduced bias in that the majority of the most cited papers report on negative impacts and the continual citation of these in the literature may have led to an overestimation of impacts.[28] Impact was most intense in areas of high elephant density, in arid savannahs and in regions in which elephants were confined by fences.[28] The analysis highlighted the difficulty of disentangling the relative contributions of climate, fire, drought, soil characteristics and herbivores to the woodland loss recorded in many of the studies – a challenge that has been emphasised elsewhere.[8,11,17]

While the damaging effects on trees are clear, elephant browsing often facilitates the establishment and spread of the shrubby and herbaceous vegetation that replaces woodland, facilitating, in turn, the biological communities associated with these alternative vegetation types.[8,10,12]

Figure 1: A bull elephant in Ntsiri Nature Reserve adjacent to Kruger National Park, South Africa. This majestic giant needs about 170 kg of green food a day and can be quite destructive in its feeding habits (photo: T. Kuiper).

The implications for biological diversity

With the effect of elephants on vegetation clearly demonstrated, what are the implications for biological diversity in general? One must evaluate the effects of elephant presence within the context of the role that large herbivores play in savannah systems. The absence of elephants and other large herbivores disrupts savannah ecosystems and may lead to bush encroachment and trophic cascades.[7,29] Given the role played by ungulates in ecosystem functioning, most notably through top-down forcing, their extirpation from an area will lead to cascading effects.[7,29]

Many studies investigating the impacts of elephants on vegetation point to the implications for biological diversity in general, but few experimentally test this assertion.[11,16,17] Greater woody plant structural diversity in the absence of elephants may or may not enhance the habitat available for a wide variety of other organisms beyond herbivores.[10,12] In Zimbabwe's Miombo woodlands, Cumming et al.[10] recorded significantly lower abundances of ants and birds in elephant-impacted woodlands compared to adjacent intact woodland. Across 30 sampling sites in the same region, Fenton et al.[19] recorded significantly lower micro-bat species richness, abundance and activity at elephant-impacted sites versus intact sites. In Amboseli National Park (Kenya), structural changes in vegetation, driven primarily by elephants, have purportedly resulted in the extirpation of both lesser kudu (*Tragelaphus imberbis*) and bushbuck (*Tragelaphus scriptus*) from the region.[18]

After 5 years of research into the ecosystem-level effects of elephants in northern Botswana, Skarpe et al.[8] concluded that they had found no ecological reason to control elephant numbers. Their results suggested neutral or even positive effects of elephant presence on other species, including small mammals and ungulates.[8] Elephant impacts were shown to increase the availability of those shrub species commonly browsed by sympatric ungulate species and the distribution of half of these species exhibited a positive correlation with elephant impact.[8] Despite an expanding elephant population, increases in both buffalo (*Syncerus caffer*) and impala (*Aepyceros melampus*) numbers were recorded during the study.[8]

In Addo Elephant National Park, South Africa, the reintroduction of apex predators (lions, *Panthera leo* and spotted hyaenas, *Crocuta crocuta*) led to the decline of small ungulate prey in an area of the park where elephant densities were high, but no such declines were recorded in an area characterised by a long history of elephant absence.[30] The authors suggested that the denser thicket of the elephant-free area provided refuge for small prey from predation by the reintroduced apex carnivores, while the fragmented thicket of the elephant-impacted area facilitated the impact of these predators on small ungulate prey.[30]

The role of human intervention

Fences and artificial water points alter elephant movements, thereby increasing the localised impact on vegetation.[22,31] Artificial water points allow greater dry season ranging, resulting in elephants being able to exploit areas they would not usually have access to during the dry season and reducing spatial refuges for vulnerable plant species.[22,31,32] Given that elephants are more faithful to dry season than wet season ranges, the dry season presence would have a more consistent impact on the landscape than would the nomadic wet season footprint.[31] Fences act as barriers to elephant movements, leading to the restriction of wet season ranges.[22,31] In addition, fences cause elephants to visit the same areas more often in the wet season, 'bunching up' against fences and causing localised vegetation impacts.[31] In combination, fences and artificial water points reduce the difference between the wet and dry season ranges of elephants and lead to vegetation impacts at a localised scale that would be avoided if elephants ranged naturally and freely.[13,31]

Conclusions

Many studies have focused on the short-term elephant impacts on a limited number of plant species, with weak potential for inference about higher-scale vegetation change which is likely modulated by other factors such as rainfall and fires. Also, change may wrongly be interpreted as undesirable when in fact it is natural. Finally, the impact of human interventions in mediating elephant impacts has to be emphasised and explored more extensively. Beyond these considerations, it is important to appreciate the integral role that elephants play in regulating savannah dynamics and to be reminded of the fact that the decline in large herbivores across the globe in recent years poses a significant threat to ecosystem integrity.

So, are elephants good or bad for biological diversity? A sweeping and simple answer to this question is elusive, in part because of the lack of research that explicitly seeks to answer it. While there is no question that elephants have driven extensive vegetation change in some areas, the consequences of these changes for community-wide biological diversity remain somewhat unclear. Whilst scientists may perceive elephants to be biodiversity thieves, this view appears to be based primarily on intuition, rather than on hard science. We found little compelling evidence for adverse effects of elephants on biological diversity.

We recommend that more rigorous and longer-term (5–15 years) experiments, designed specifically to test the impacts of elephants on biodiversity, be the focus of future research. Studies that compare a variety of robust indices of community-level biological diversity between elephant-impacted and control areas across multiple sites and over broad spatial scales are necessary before any concrete patterns become evident. Until then, conservationists will do well not to cry elephant.

Acknowledgements

This commentary originated from an essay assignment, set by Dr Daniel Parker and written by Timothy Kuiper, which formed part of the Zoology Honours course at Rhodes University (2013). We are thankful to Jaap Kuiper, Martin Villet, Adrian Craig and Jessica Comly for comments on earlier drafts.

References

1. Sutherland WJ, Bardsley S, Clout M, Depledge MH, Dicks LV, Fellman L, et al. A horizon scan of global conservation issues for 2013. Trends Ecol Evol. 2012;28(1):16–22. http://dx.doi.org/10.1016/j.tree.2012.10.022
2. Brooks TM, Mittermeier RA, Da Fonseca GA, Gerlach J, Hoffmann M, Lamoreux JF, et al. Global biodiversity conservation priorities. Science. 2006;313(5783):58–61. http://dx.doi.org/10.1126/science.1127609
3. Noss RF. Indicators for monitoring biodiversity: A hierarchical approach. Conserv Biol. 1990;4(4):355–364. http://dx.doi.org/10.1111/j.1523-1739.1990.tb00309.x
4. Groom MJ, Meffe GK, Carroll CR. Principles of conservation biology. Sunderland: Sinauer Associates; 2006.
5. Whyte IJ, Van Aarde RJ, Pimm SL. Kruger's elephant population: Its size and consequences for ecosystem heterogeneity. In: Du Toit JT, Biggs H, Rodgers K, editors. The Kruger experience: Ecology and management of savanna heterogeneity. Washington DC: Island Press; 2003. p. 332–348.
6. Hobbs RJ, Harris JA. Restoration ecology: Repairing the earth's ecosystems in the new millennium. Rest Ecol. 2001;9(2):239–246. http://dx.doi.org/10.1046/j.1526-100x.2001.009002239.x
7. Van Langevelde F, Van De Vijver CADM, Kumar L, Van De Koppel J, De Ridder N, Van Andel J, et al. Effects of fire and herbivory on the stability of savanna ecosystems. Ecology. 2003;84(2):337–350. http://dx.doi.org/10.1890/0012-9658(2003)084[0337:EOFAHO]2.0.CO;2
8. Skarpe C, Aarrestad PA, Andreassen HP, Dhillion SS, Dimakatso T, Du Toit JT, et al. The return of the giants: Ecological effects of an increasing elephant population. AMBIO. 2004;33(6):276–282.
9. Asner GP, Levick SR, Kennedy-Bowdoin T, Knapp DE, Emerson R, Jacobson J, et al. Large-scale impacts of herbivores on the structural diversity of African savannas. Proc Natl Acad Sci USA. 2009;106(12):4947–4952. http://dx.doi.org/10.1073/pnas.0810637106
10. Cumming D, Fenton M, Rautenbach I, Taylor R, Cumming G, Cumming M, et al. Elephants, woodlands and biodiversity in southern Africa. S Afr J Sci. 1997;93(5):231–236.
11. Western D, Maitumo D. Woodland loss and restoration in a savanna park: A 20-year experiment. Afr J Ecol. 2004;42(2):111–121. http://dx.doi.org/10.1111/j.1365-2028.2004.00506.x
12. Asner GP, Levick SR. Landscape-scale effects of herbivores on treefall in African savannas. Ecol Lett. 2012;15(11):1211–1217. http://dx.doi.org/10.1111/j.1461-0248.2012.01842.x
13. Van Aarde RJ, Jackson TP. Megaparks for metapopulations: Addressing the causes of locally high elephant numbers in southern Africa. Biol Conserv. 2007;134(3):289–297. http://dx.doi.org/10.1016/j.biocon.2006.08.027
14. Leuthold W. Changes in tree populations of Tsavo East National Park, Kenya. Afr J Ecol. 1977;15(1):61–69. http://dx.doi.org/10.1111/j.1365-2028.1977.tb00379.x
15. Guldemond R, Van Aarde R. The impact of elephants on plants and their community variables in South Africa's Maputaland. Afr J Ecol. 2007;45(3):327–335. http://dx.doi.org/10.1111/j.1365-2028.2007.00714.x
16. Barnes R. Effects of elephant browsing on woodlands in a Tanzanian National Park: Measurements, models and management. J Appl Ecol. 1983;20(2):521–539. http://dx.doi.org/10.2307/2403524
17. Ben-Shahar R. Woodland dynamics under the influence of elephants and fire in northern Botswana. Vegetation.1996;123(2):153–163. http://dx.doi.org/10.1007/BF00118268
18. Western D, Gichohi H. Segregation effects and the impoverishment of savanna parks: The case for ecosystem viability analysis. Afr J Ecol. 1993;31(4):269–281. http://dx.doi.org/10.1111/j.1365-2028.1993.tb00541.x
19. Fenton MB, Cumming DH, Rautenbach I, Cumming GS, Cumming MS, Ford G, et al. Bats and the loss of tree canopy in African woodlands. Conserv Biol. 1998;12(2):399–407. http://dx.doi.org/10.1046/j.1523-1739.1998.96376.x
20. Lewin R. In ecology, change brings stability. Science. 1986;234:1071–1073. http://dx.doi.org/10.1126/science.234.4780.1071
21. Parker D, Bernard R. Lessons from aloes in the Thicket Biome: Reconstructing past elephant browsing to understand the present. S Afr J Sci. 2008;104(5–6):163–164.
22. O'Connor TG, Goodman PS, Clegg B. A functional hypothesis of the threat of local extirpation of woody plant species by elephant in Africa. Biol Conserv. 2007;136(3):329–345. http://dx.doi.org/10.1016/j.biocon.2006.12.014
23. Campbell AC. History of elephants in Botswana. In: Hancock P, Cantrell M, Hughes S, editors. The future of Botswana's elephants. Gaborone: Kalahari Conservation Society; 1990. p. 5–15.
24. Colegrave R, Lungu E, Muwezwa M. What is happening to the elephants in Botswana? Revue de Bio Matematique. 1992;30:49–55.
25. Junker J, Van Aarde RJ, Ferreira SM. Temporal trends in elephant *Loxodonta africana* numbers and densities in northern Botswana: Is the population really increasing? Oryx. 2008;42(01):58–65. http://dx.doi.org/10.1017/S0030605308000756
26. Owen-Smith N, Kerley G, Page B, Slotow R, Van Aarde RJ. A scientific perspective on the management of elephants in the Kruger National Park and elsewhere. S Afr J Sci. 2006;102:389–394.
27. Bernard RTF, Parker DM. The use of archaeological and ethnographical information to supplement the historical record of the distribution of large mammalian herbivores. S Afr J Sci. 2006;102:117–119.
28. Guldemond R, Van Aarde RJ. A meta-analysis of the impact of African elephants on savanna vegetation. J Wildlife Manage. 2008;72(4):892–899. http://dx.doi.org/10.2193/2007-072
29. Pringle R, Young T, Rubenstein D, McCauley D. Herbivore-initiated interaction cascades and their modulation by productivity in an African savanna. Proc Natl Acad Sci USA. 2007;104:193–197. http://dx.doi.org/10.1073/pnas.0609840104
30. Tambling CJ, Minnie L, Adendorff J, Kerley GIH. Elephants facilitate impact of large predators on small ungulate prey species. Basic Appl Ecol. 2013;14(8):694–701. http://dx.doi.org/10.1016/j.baae.2013.09.010
31. Loarie SR, Aarde RJV, Pimm SL. Fences and artificial water affect African savannah elephant movement patterns. Biol Conserv. 2009;142(12):3086–3098. http://dx.doi.org/10.1016/j.biocon.2009.08.008
32. De Beer Y, Van Aarde R. Do landscape heterogeneity and water distribution explain aspects of elephant home range in southern Africa's arid savannas? J Arid Environ. 2008;72(11):2017–2025.

Taking the transformation discourse forward: A response to Cloete, Dunne and Moultrie and Dorrington

AUTHORS:
Kesh S. Govinder[1]
Nombuso P. Zondo[1]
Malegapuru W. Makgoba[2]

AFFILIATIONS:
[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa
[2]Office of the Vice-Chancellor, University of KwaZulu-Natal, Durban, South Africa

CORRESPONDENCE TO:
Kesh Govinder

EMAIL:
Govinder@ukzn.ac.za

POSTAL ADDRESS:
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X54001, Durban 4000, South Africa

KEYWORDS:
transformation; universities; demographics; equity index; ranking

HOW TO CITE:
Govinder KS, Zondo NP, Makgoba MW. Taking the transformation discourse forward: A response to Cloete, Dunne and Moultrie and Dorrington. S Afr J Sci. 2014;110(3/4), Art. #a0060, 8 pages. http://dx.doi.org/10.1590/sajs.2014/a0060

We respond to the recent commentaries[1-3] on our articles[4,5]. We indicate how statements have been misread, misinterpreted or viewed in the narrowest manner possible. Any suggestion of bias or ‘smoke and mirrors’ in our article[5] is rejected. On the contrary, we emphasise that our approach was transparent and provided a snapshot of the sector in an objective manner.

Introduction

We noted the very passionate commentaries[1-3] on our articles[4,5] that have recently appeared in the *South African Journal of Science*. It is clear that all these authors take the issue of transformation in South Africa and quality high-level knowledge production of higher education institutions very seriously and we laud their commitment to this important cause. It is a great pity though, that this passion has led to an incomplete reading of our work and/or the most narrow interpretation possible of our results and conclusions. In particular, several statements are taken out of context, while others are simply misinterpreted; still others indicate a lack of understanding of the South African context of the study (for example, Dunne’s[2] criticism of the term ‘equity’ is puzzling). In addition, we believe that statements like the applicability of the formula is ‘morally dubious’[2] or ‘questionable mission of the paper’[3] (and many others that appear throughout the commentaries of Dunne[2] and Moultrie and Dorrington[3]) are, at best, sound bites not befitting a scholarly critique of a peer-reviewed paper. What we find especially surprising is that none of these emotive statements are followed up with clear, substantiated evidence to support the sentiments espoused.

We must emphasise an important point that has continually been misinterpreted and misrepresented – our Equity Index (EI) provides the first quantitative *measure* of transformation. This should not be interpreted to mean that our Equity Index is the final word on transformation. It should also not be taken to mean (rather mischievously, in our opinion) that transformation is only about numbers. As stated in our paper[5]:

> *In the South African context, transformation refers more specifically to change that addresses the imbalances of the past (apartheid) era. It has many facets, including demographic and systemic change. However, regardless of the different components and qualitative measures for transformation, the ultimate (and most important) indicator is that of demographics (racial and gender statistics).*

In measuring the distance between an organisation’s demographics and the national demographics, one can conclude how effective that organisation’s transformation activities are. We do not pass judgement on what those transformation activities are or indeed what they should be. The EI merely indicates how effective an organisation has been (via its own transformation processes and policies) in ensuring that it reflects the relevant demographics and the constitutional imperative of transformation.

Context and data

Any scientific study must be read within its own context (i.e. within the parameters indicated at the outset). For our study, we were clear about the parameters: the study was undertaken using publicly available data and based on the Constitution, White Paper 3 of 1997 and the equity laws on transformation of higher education. It must be stated that *it is unconstitutional for any organisation not to transform*. The first set of data was obtained from the most recent census and sourced from official documentation released by StatsSA[6]. Such data may indeed be open to criticism and revision from time to time. Notwithstanding such possible future revisions, we undertook our study using the current official demographic statistics available. We have already justified[5] our use of national versus regional demographics. Notwithstanding this justification, we note that the results will change significantly only for a few institutions when regional as opposed to national demographics are used, but our findings and conclusions for the sector will remain unchanged.

The second set of data was the HEMIS data provided by the Department of Higher Education and Training (DHET)[7]. Again, this set of data is official data that was compiled by the institutions themselves in terms of race and gender. As such, this data may indeed contain errors of numbers and of definitions and be open to debate. As an example, consider the definition of Instructional/Research Staff which lends itself to many differing interpretations. The universal gold standard of Instructional/Research Staff is the holding of a PhD. If this universal benchmark was to be applied (especially with the distortions of our colonial-apartheid legacy) one can foresee the obvious problem – many academic staff in Medicine, Law, Commerce/Accountancy, Engineering and Computer Science would be removed from consideration as those professions tend not to require PhDs. It is therefore clear that defining such benchmarks is not without controversy. Notwithstanding this challenge, the data, as they stand, are still the official data maintained by the Department and we used this data at face value.

Finally, the research output of each institution is determined by DHET[8] based upon audited submissions by the different institutions. This output influences (volume as opposed to efficiency) the calculation of the research component of the block grant provided by the Department and so is crucial for each institution in the country. It is unfortunately true that this data equates all publications in ISI journals, regardless of the quality of the journal (although, even here, there is no universally accepted indicator of journal quality – even impact factors are debatable). Again, this is the official national data available and we used this, as generated, for our study.

However, we did analyse the national research output in two ways: Total Research Output (Volume) vs Equity Index and Per Capita Research Output (Efficiency) vs Equity Index. Both volume of research output and research efficiency are important factors to consider.

It would seem to be self-evident, but we feel that we have to emphasise that any study and its conclusions must be read in its proper context and with an understanding of its parameters. We believe that such an approach is in line with proper scientific discourse.[9]

The study by Dorrington et al.[9] is illustrative of this approach. They studied projections of mortality (AIDS deaths) in South Africa, and compared four models: ASSA600, the US Bureau of the Census, the United Nations and the Metropolitan–Doyle. The projections for deaths due to AIDS differed significantly between the US Bureau of the Census and the Metropolitan–Doyle Model (by approximately 86% in fact). When they recast the balance of equations to ASSA600 (their model) base population and migration, the variation reduced to 82% and stood at 36% between ASSA600 and the US Bureau of the Census. Rather interestingly, the report describes the results of the other models using phrases like 'appear low' or 'appear high'. None of those results were said to be 'wrong' or 'flawed'. Presumably it was understood that these different models used different assumptions to arrive at these significantly different results within the same population during the same period of study. Importantly, one common finding of all the models was that AIDS deaths increased significantly. This conclusion was not disputed regardless of the fact that each model incorporated different assumptions. We would expect the same standard to be applied to our work, or indeed any other scientific study.

We do acknowledge that the issue of inclusion or exclusion of foreigners has not been settled. Initially, our study included foreigners in the different categories. However, as a result of concerns raised by a reviewer, foreign students were removed from the data. However, as the original data aggregated all the foreigners together for staff, the EIs as presented in our staff analysis[5] does indeed include them. This is important as it ensures that both the divisors and numerators in the equity-weighted total research output and per capita research output are consistent. Without including foreigners into the calculations, these indices would be meaningless. Notwithstanding this, we acknowledge that there is no agreement on if and how non-South Africans should be included in this study (or indeed in any discussion incorporating demographics in South Africa). Indeed, as indicated in our paper, the inclusion of these individuals into the student data did improve EIs.

As those debates on the inclusion or exclusion of foreigners continue and are eventually settled, the benchmarks can be easily modified.

It must be pointed out that the set of data points we used was six dimensional (although there is a relationship between the coordinates as the sum of the first four components cannot exceed 100% and the sum of the final two components cannot exceed 100% either) while Moultrie and Dorrington[3] prefer an eight-dimensional set (with the sum of the eight components not exceeding 100%). These are two independent sets and results obtained in one set cannot be extrapolated to the other. The advantage of our approach is that one can actually determine a race EI (using data points with only four (race) components), EI_r, a gender EI (using data points with only two (gender) components), EI_g, and an overall EI which is a combination of these two previous EIs via $EI = \sqrt{EI_r^2 + EI_g^2}$. It was this aggregate combination possibility that led us to use $n=6$. Importantly, the EI_r and EI_g both form subspaces of the EI space using this approach. For $n=8$, we do not have this mathematical structure – one cannot find EI_r and EI_g directly in that space.

However, this does not mean that we ignored the $n=8$ case. In fact, all EIs were calculated using $n=8$ as well. Given the above characteristics of the $n=6$ data, we preferred to publish those results as opposed to the $n=8$ data. We present those results in Table 1 and Table 2 as well as Figure 1 and Figure 2. It may appear that this leads to lower EIs overall but, as indicated earlier, one cannot compare the $n=6$ results to the $n=8$ results as the reference points are different (and lie in different subspaces – six dimensional in the first case as opposed to eight in the latter case). While the ranks of some institutions may have changed marginally (but not significantly), the overall conclusions are similar to those presented in our paper[5] (see later for a specific comparison with the results of Moultrie and Dorrington[3]). In summary, using $n=6$ is useful as a combination of the race and gender EIs while using $n=8$ is more useful for exclusively overall EIs. Neither can be said to be wrong, but each must be considered in its own context and they cannot be compared.

Table 1: Equity Indices (EIs) for South African student enrolment and graduation at South African universities ($n=8$)

Institution	Enrolment EI	Ranking	Graduate EI	Ranking	Equity Efficiency Index
Cape Peninsula University of Technology	28.8	18	33.3	17	-4.5
Central University of Technology	7	1	5.9	1	1.1
Durban University of Technology	11.2	4	13.8	4	-2.6
Mangosuthu University of Technology	17	9	17.3	7	-0.3
Nelson Mandela Metropolitan University	19.9	12	25.5	12	-5.6
North-West University	25.4	17	27.6	15	-2.2
Rhodes University	38.8	20	41.9	20	-3.1
Tshwane University of Technology	10.5	3	11.3	2	-0.8
University of Cape Town	44.8	22	52.3	22	-7.5
University of Fort Hare	15.4	6	17.7	8	-2.3
University of Johannesburg	8	2	15.2	5	-7.2
University of KwaZulu-Natal	24	14	26.9	14	-2.9
University of Limpopo	15.9	8	16.7	6	-0.8
University of Pretoria	32.9	19	36.3	18	-3.4
University of South Africa	15.4	6	24.1	11	-8.7
University of Stellenbosch	66	23	66.1	23	-0.1
University of the Free State	21.7	13	38.3	19	-16.6
University of Venda	17.6	10	18.8	9	-1.2
University of Western Cape	43.8	21	44.2	21	-0.4
University of the Witwatersrand	24.3	15	29.9	16	-5.6
University of Zululand	25	16	26.7	13	-1.7
Vaal University of Technology	13.5	5	13.3	3	0.2
Walter Sisulu University	18.3	11	20	10	-1.7

Table 2: Equity Indices for staff categories at universities in South Africa together with institutional rankings within each category (*n*=8)

Institution	Overall	Rank	Exec	Rank	Instruct	Rank	NP Admin	Rank	Service	Rank	Spec	Rank	Technical	Rank	Crafts	Rank
Cape Peninsula University of Technology	45.2	16	44	13	49.5	13	45.7	15	65.6	20	44.3	14	51.8	18	104.4	22
Central University of Technology	31.1	11	37.6	10	46.6	10	43.4	14	12.3	2	29	8	39.2	10	54.7	10
Durban University of Technology	40	15	58.9	17	47.2	11	34.9	11	34.9	16	44.2	13	50.9	17	77.7	18
Mangosuthu University of Technology	12.5	5	21.9	4	25.2	5	20.8	5	19.2	8	12.6	2	23.1	3	65	14
Nelson Mandela Metropolitan University	49	17	55.7	15	64.6	20	49.4	17	17.9	5	50.1	17	41.6	11	53.4	8
North-West University	51.1	20	69.8	23	59.4	17	65.9	22	13.5	4	65.3	21	50.6	16	60.2	13
Rhodes University	30.5	10	65.2	19	69	22	47.7	16	12.9	3	52.7	18	47.4	15	57.9	11
Tshwane University of Technology	22.6	8	29.2	6	37	8	17.7	4	21	10	33	9	27.7	5	57.9	11
University of Cape Town	54.1	21	62.5	18	63.4	18	58.2	19	54.8	19	56.9	20	60.9	20	87.9	20
University of Fort Hare	9	1	26.2	5	22.6	4	17.3	3	67.6	21	11.4	1	19.9	2	74.8	16
University of Johannesburg	34.9	14	55.1	14	53.5	15	38.4	13	39.4	18	44.9	15	35.7	9	53.2	7
University of KwaZulu-Natal	34.1	13	38.9	11	49.3	12	34.9	11	34.2	15	42.4	12	23.6	4	88.2	21
University of Limpopo	10.6	3	12.9	1	17.4	2	16	2	21.1	11	15	3	33.2	7	47.6	5
University of Pretoria	49.2	18	67.6	21	65.7	21	49.8	18	31	13	45.1	16	59.3	19	54	9
University of South Africa	27.1	9	36.4	8	41.6	9	21.4	6	31.4	14	35.3	10	45.3	14	44.8	4
University of Stellenbosch	64.9	23	67.7	22	72.3	23	69.3	23	73.3	23	69.6	22	66.2	22	80.5	19
University of the Free State	50.5	19	67.1	20	64.1	19	59.6	20	9.8	1	78.8	23	63.8	21	70.1	15
University of Venda	16.6	7	15.9	2	26.3	6	22.5	7	20.9	9	20.8	4	44.2	13	31.5	2
University of Western Cape	56.6	22	56.6	16	50.2	14	65.8	21	72.1	22	54.2	19	72.5	23	110.5	23
University of the Witwatersrand	32.3	12	42.4	12	56.6	16	29.6	9	35.1	17	35.8	11	34.4	8	32.8	3
University of Zululand	12.2	4	29.4	7	18.8	3	12.3	1	22.9	12	21.1	5	32	6	74.8	16
Vaal University of Technology	15.5	6	36.9	9	33.4	7	32	10	18.1	6	26.9	7	17.5	1	51.7	6
Walter Sisulu University	9.6	2	19.7	3	11.6	1	25.3	8	18.7	7	23.4	6	42.8	12	30.1	1

Exec, executive/administrative/managerial professionals; Instruct, instructional/research professionals; NP admin, non-professional administrators; Service, service staff; Spec, specialist/support professionals; Technical, technical staff; Crafts, crafts/trade staff.

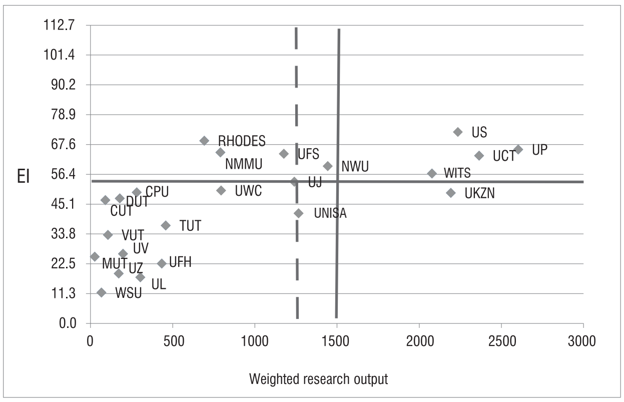


Figure 1: The 2011 total weighted research productivity versus the instructional/research professional staff Equity Index (*n*=8).

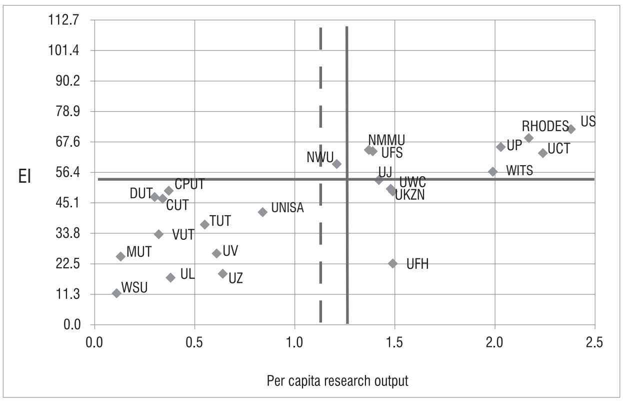


Figure 2: The 2011 per capita research output versus the instructional/research professional staff Equity Index (*n*=8).

Distance formula

The racial and gender demographics in South Africa are reported in terms of percentages. As the percentages are reported in different categories, it is not easy to determine whether one is moving towards a particular benchmark or not. Take for example, the national racial percentages of (79.2, 8.9, 2.5, 8.9). If an organisation moves closer to these percentages in the first and second categories, but further away from the percentages in the third and fourth categories, it is not obvious whether progress has been made towards reflecting those percentages. Because we are concerned with moving closer or further away from a particular point, it would seem that the distance formula is a natural way of determining this.

In general, we are concerned with a set consisting of points with components consisting of demographic percentages. The distance between any two points in a particular set is given by a metric. In the case of Euclidean (flat) space, the metric reduces to the generalisation of the Theorem of Pythagoras in *n* dimensions. Then, the distance formula we use is given by

$$distance = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

where the x_i's and the y_i's are components of the two points $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$. This translated naturally into the Equity Index[4,5] given as

$$Equity\ Index = \sqrt{\sum_{i=1}^{n} (org_i - demdat_i)^2}$$

where org_i refers to components of the organisation's demographic data point and $demdat_i$ refers to the components of the benchmark demographic data point. In our opinion, this is the simplest *form* of the formula. Other, more complicated forms exist, but they do not change the result of the calculation. We note that Moultrie and Dorrington[3] preferred the form

$$EI = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (p_{ij} - p_{ij}^{b})^2}$$

However, the problem with the above formula is that it can only be used to calculate the EI if each data point has components that are disaggregated in terms of gender and race (for a total of eight components – in fact, their definition of p_{ij} necessitates $n=4$ and $m=2$). In the form that we use, one can utilise eight components as used by Moultrie and Dorrington[3] or six components as used by us[5].

Unlike the usual distance between two points in which the components of each point are independent, we use demographic data points in which the components add up to 100% (in the case of eight demographic categories) or 200% (in the case of six demographic categories – which was the case for our study). As a result, when trying to optimise the path for an organisation's demographics to reflect the benchmark, one has to be a little careful; one has to undertake a constrained optimisation (technically, optimisation on a lower dimensional hypersurface – a paraboloid of lower dimension than the original hypersurface). Thus, there is at least one degree of freedom less than is ordinarily the case. We have used this non-independence of coordinates to good effect by determining a maximum EI. Usually, one would not think of a maximum distance from a reference point for which the components are completely unconnected. However, because of the nature of the components of the points we are utilising, we *can* talk of a maximum. We do this by taking a 'worst case scenario' of only Indian males comprising the demographics (this is done as Indians comprise the lowest racial percentage and males the lowest gender percentage). Calculating the distance of this point ((0,0,100,0,0,100) for $n=6$ or (0,0,0,0,0,100,0,0) for $n=8$) from the relevant benchmark gives the maximum EI. Subsequently, this maximum was divided into five equal parts called quintiles. (This terminology is perhaps unfortunate as we are not referring to partitioning into quantiles as a statistician may assume, but simply the division of a particular interval into five equal parts.) This afforded an organisation the benefit of monitoring its demographic progress towards the benchmark in terms of progression to new quintiles. Perhaps paradoxically, we spoke of a sixth quintile, the *zeroth* quintile, to indicate a region of tolerance. We admit that this use is rather quaint, but thought that it did speak directly to what we were illustrating.

We have applied the distance formula as it was originally described in a straightforward, transparent manner. There are few formulae that are indeed so simple to use. We cannot understand how and why this constitutes 'fumbling mathematical conjuring'[2].

Direct responses

In what follows, we refer to the Commentary of Cloete[1] as NC, that of Dunne[2] as TD and that of Moultrie and Dorrington[3] as MD. Quotes from these commentaries are indented while quotes from our paper are included as part of the narrative within quotation marks.

Unlike, and in complete contradiction to the Commentaries of Dunne and Moultrie and Dorrington, Cloete's Commentary sees *value* in the Equity Index study. He starts off by reflecting on a draft of our paper that he had seen:

> NC: *I noted when I read the draft of Govinder et al.'s paper on equity indices that it equated equity with transformation, and delinked equity from development and performance. This draft version of the paper fell into the trap of a prevailing South African condition: using transformation as a code word for race. Furthermore, the formula used in the paper produced a result in which several of the most equitable institutions were those being run by a government-appointed administrator. By this, the authors implied their promotion of high equity, yet also regarded the existence of dysfunctional institutions as a given in their proposed model for the South African university system.*

Thereafter, he goes on to indicate how our published paper addressed these two crucial issues:

> NC: *The paper on equity indices, published in the* South African Journal of Science, *certainly responds to both these criticisms. Firstly, equity is used mostly in reference to the formula as described in the paper, although the focus of equity is racial, being mainly African. Secondly, a serious attempt is made to reconcile the well-known Harold Wolpe tension between equity and development, as described by Cloete et al. While I will argue that the attempt is not entirely successful, the approach of developing empirical indicators to reflect the equity–development duality of transformation is to be lauded as it is a step towards developing South African indicator-based performance clustering systems. My time spent at the Shanghai Jiao Tong University's Centre for World Class Universities during early November 2013 has made it even clearer to me that, while for the foreseeable future the Jiao Tong type of methodology will continue to make a considerable contribution to debate and controversy, it will not assist much in strengthening universities in Africa.*

> *Govinder et al. are correct when they assert that equity-weighted research output goes beyond the Centre for Higher Education Transformation (CHET) clusters, which were based mainly on performance and efficiency in knowledge production. The more recent CHET clustering 3 in 2010 has been expanded to include factors such as staff qualifications, undergraduate-to-masters graduation rates and high-level knowledge production (doctorates and research publications). This latest CHET clustering has shown that, in addition to those usually in the top group of higher education institutions (such as UCT, Stellenbosch University and Wits University), some 'on-the-move' institutions, such as UKZN, North-West University and the University of the Western Cape, have moved into the top group.*
>
> *Govinder et al. are also correct in pointing out that some of their results do not square up with the CHET differentiated clusters. For example, their high rating for Unisa – in terms of both the graduation Equity Index and the weighted research output – is completely contradictory to the performance of Unisa in the South African system as shown by CHET. Similarly, their low ranking for Rhodes University is contradictory to the CHET finding that Rhodes is one of the three most efficient knowledge producers in terms of weighted publication per staff member. It appears that by not using staff: research ratios, the Equity Index formula has skewed results in favour of larger institutions.*

We have quoted Cloete[1] in full so as not to distort, misrepresent or take out of context any statements in his Commentary. Firstly, Cloete is generally correct in his interpretation of our work in this part of his Commentary. Cloete then devoted most of his Commentary on his experiences in the National Commission on Higher Education, the nature of debates therein and the many uncomfortable issues they confronted but never resolved nor settled. They are important and give an interesting context to our study.

He then makes the following three major criticisms of our article: That our study

> NC: *implied their promotion of high equity, yet also regarded the existence of dysfunctional institutions as a given in their proposed model for the South African university system.*

In fact, to the contrary, as stated in our article[5]: "The South African dilemma is that some of the institutions with good equity are poor knowledge producers, and vice versa. In this study, a group of 8–11 universities (Figure 3) with very good EIs but with very low total weighted research productivity outputs is discernible (a similar number have good EIs but low per capita research output). Even with their good EIs, there is no improvement in their equity-related ranking in high-level knowledge production. This finding suggests that the quality of equity transformation is essential in knowledge-producing organisations such as universities. This group constitutes an example of how equity transformation without quality leads to unintended negative consequences. As the status quo, this group adds no value to national development, which the Council on Higher Education aptly put as follows:

> *High quality higher education is crucial for social equity, economic and social development and a vibrant democracy and civil society. If higher education does not produce knowledgeable, competent and skilled graduates, generating research and knowledge, and undertaking responsive community service, then equity, development and democracy will all be constrained."*

and

"The study has shown the general applicability of the formula and emphasised the essential role of high-level knowledge production in the quality of equity during the transformation process."

Further,

> NC: *In terms of the nature of higher education, there appears to be a fundamental flaw in the Equity Index assumption that the university should be a mirror of national demographics. The university is a specific institution in society that is supposed to lead rather than reflect society. A forthcoming book by Castells and Himanen, in discussing Amartya Sen's* Development as Freedom *(1999) and John Rawls'* A Theory of Justice *(1971), highlights the argument that, while all citizens are equal before the law and are all entitled to dignity, this is not the case in terms of capability, particularly if capability is understood as performance rather than potential.*

A gifted child in a family performs better than his/her siblings but is nevertheless a mirror and reflective of that family. Indeed universities reflect a different type of people in a society. However, universities exist within particular societies, with laws, values and cultures leading and serving those societies. This close interlinkage was recognised early – in 1155 Emperor Frederick Barbarossa said the twin role of a university was to produce rational knowledge that would 'illuminate the world and maintain social order',[10] i.e. to serve the intellect and society. The production of rational knowledge often arises out of problems within society. Universities could not do this without being mirrors and without mirroring society. There are no universities existing out of societies or for themselves in society. Take the UK, the Indian or the Chinese university systems. Each one of these societies is reflected in their university systems in terms of students and staff profile capabilities. The overwhelming majority of staff and students in the UK, Indian and Chinese university systems are UK, Indian and Chinese nationals. Furthermore, the following examples illustrate the point of capable citizens and their societies even better[11]:

1. Louis Pasteur's – a French citizen – fascination with micro-organisms, for example, led him down the applied path towards understanding the pathology of disease, the creation of alcohol and commercially viable high-quality vinegar. Pasteur developed his science by accepting problems presented by a Lille industrialist, the Ministry of Agriculture and Napoleon III.
2. Lord Kelvin's – a British citizen – physics was inspired by a deeply industrial view and the needs of the British Empire.
3. Keynes' – a British citizen – magna opera *A Treatise on Money* and *The General Theory of Employment, Interest and Money* were inspired and written in the context of massive unemployment and the grinding misery of economic depression during the 1930s.

So, the notion that because universities play a particular role in society (leading society) and are inhabited by scholars and students of particular capabilities is not incompatible with reality and the understanding that these institutions and their special inhabitants are representative, reflective and reflected of these societies. This capability representativity occurs even today despite the idea of internationalisation. Internationalisation occurs within the constraints of nations, their constitutions, their laws and values. International students are often required to return home to serve their nations or societies. For example our own Constitution, equity laws and White Paper 3 of 1997 (which anchors our Equity Index study) enjoin our universities to play a leading role in transformation but also to function in particular ways to create a representative, equitable and high-quality 'single coordinated national higher education system' that is transforming and reflective of the diversity of the South African nation. The South African taxpayer funds and invests in the system to build South Africa and not a nation 'in Esperanto', without an identity. The National Commission on Higher Education was there to develop and lay the foundations of a first rate transformed South African higher education system of quality within the constraints of the nation's Constitution.

Finally,

> NC: *Another assumption underpinning the arguments of the authors is that the slow progress is as a result of a lack of institutional compliance. Not once is the question raised as to the role of the national Department of Higher Education and Training and its contribution to the problem.*

While we did not use the term 'national Department of Higher Education and Training' we did use the word 'government' instead. Specifically: "Given the vast (government) investment in higher education since 1994 (over R236 billion up to and including 2013), the equity returns need to be interrogated. This study shows that it is difficult to transform 'privilege' voluntarily and suggests that extraordinary measures are needed. …The question remains as to the reasons behind this slow progress: Is it passive resistance or a denial of failure by the sector? Is it the abuse of autonomy or an abhorrence of accountability by the sector? Has government failed to provide clear unambiguous steering or monitoring mechanisms or has it been cowed by the voice of the 'privileged' at the expense of the voice of the disadvantaged majority, shying away from doing that which is commonsense in a democracy?"

and

"It may become necessary for government to set knowledge production targets or set this group on a different mission/trajectory to contribute to national development". The word 'government' appears seven times in our study!

Cloete[1] then concludes as follows:

> NC: *The task is thus to build a new post-secondary differentiated higher education system with built-in quality checks. This system should include a mix of research-led universities, universities that are mainly undergraduate teaching institutions, a further education and training college sector that is mainly post-matric and vocationally orientated, and a private sector that is market driven.*

Our comment was "Regardless of whether one focuses on the CHET clusters or our groupings, it is clear that, for transformation to advance and succeed, government has to address differentiation urgently: firstly, in the staff and student composition of institutions and, secondly, in their performance with respect to research productivity".

and

"In terms of differentiation, not all universities have to be high-level knowledge producers, but if they are to contribute to development they at least have to provide a quality undergraduate education and improve their very poor throughput rates. It does not help that they take in the 'disadvantaged' and then do not add value to their skills and certification. It may become necessary for government to set knowledge production targets or set this group on a different mission/trajectory to contribute to national development. On the other hand, the high-level knowledge producers with poor EIs need to be set equity targets, which could be rather complex but customised".

We now focus on some specific comments of the other two commentaries. We do not address every single comment as that would mean repeating almost our entire article verbatim.

> MD: *Despite the authors' assertion that either method of calculating the index 'is relevant', their approach double counts and is not mathematically correct. Correcting for this error has the effect of reducing all the EIs by approximately 30%. If the primary concern is simply the ranking of the universities (as opposed to drawing any other conclusions) the correction does not change those rankings, but it does impact, inter alia, the maximum determined values of the EI; the relationship between research output per capita and EI; and the demarcation of the quadrants, and hence the conclusions drawn from partitioning the plot area into quadrants.*

We were completely transparent about the fact that we used $n=6$ in our study and have already indicated above the usefulness of each calculation. Interestingly, by using $n=8$ and modifying the data as they believe is necessary for a 'correct' analysis, Moultrie and Dorrington[3] obtain results completely in accordance with our results. If we look at their Figure 2[3] and interchange axes, we obtain our Figure 4[5]. The actual values are different, but the relative positions of all the data points are unchanged. Indeed, as stated in our work, the positions of UKZN and UWC on the figures are tenuous (as evidenced by the statements "UKZN has a good EI and high productivity (although it should still address its EI)" and "UWC and UKZN (both of which just enter this quadrant)") – they are subject to change with slight perturbations of the data. This was borne out by the changes to the data incorporated by Moultrie and Dorrington[3].

> MD: *They treat the EIs as additive (they are not) and thus produce what they term an 'equity efficiency index' to measure change by differencing the indices of enrolled students and graduates; or, even more absurdly, they sum the EIs for the various categories of employees (presented in Figure 2 of their paper) to compare universities.*

Firstly, as the EIs are calculated using the distance formula, mathematically they can indeed be added together or subtracted from each other (as can distances in general). It is the interpretation of this addition that is important. In subtracting the EIs of enrolled students and graduates, we have a sense of whether an institution's equity profile has moved away from or towards the national demographics. (We acknowledge that this gives a snapshot of the situation for 2011. More data points will allow a proper comparison of the results for a particular cohort.) Further, it must be noted that Figure 2 is a graphical representation of the data in Table 5. A simple additive test will indicate that the different EIs in Table 5 are *not* added together. In Figure 2, the EIs are presented together for illustrative comparison purposes of each category and nothing else (although we acknowledge that the greyscale figure makes this less illustrative than we would like).

> MD: *The absurdity is compounded by the small numbers of staff in several categories – a problem of which they are aware, but appear to ignore in this instance. For example, they calculate an EI for the 'crafts and trades' group across the 23 HEIs, despite there being fewer than 20 such staff at 13 of the 23 institutions.*

We have already mentioned in the paper: "(We note that the actual number of staff in the category 'crafts/trade' is quite small. We have previously cautioned against using an EI for small numbers as the EI can change dramatically with a small change in individual staff employed.)" The main reason for including this data (and ranking) is for completeness and transparency. We wanted to include all HEMIS categories as provided by DHET. In addition, this category is the one in which UKZN performs the worst. In order to show that the report is unbiased we felt the need to include this information. Unfortunately, this aim was not achieved as there are various suggestions that we presented the report in order to show UKZN in the best possible light and thus the mission of the paper was 'questionable' and 'morally dubious'.

This latter implied bias on our part is also apparent in:

> TD: *Indeed, one of the hypotheses offered by the authors is that several universities (other than their own university – the University of KwaZulu-Natal, UKZN) are currently impervious to equity objectives.*

Contrast this with "every institution analysed has a challenge with respect to some EI." Further, the n=8 results presented in Table 1 and Table 2 as well as in Figure 1 and Figure 2 show either unchanged or improved results for UKZN.

> MD: *Apart from the fact that they do not standardise for the size of the institution…*

On the contrary: "In order to address the issue of institutional 'size', one could also look at research in terms of per capita output. We present this perspective in Figure 4. The quadrant boundaries have been determined in the same manner as for Figure 3. Here we note, very pleasingly, that no university falls firmly into the top left quadrant (which indicates poor EIs and low per capita research output) as all of these universities have moved across into the top right quadrant (with still poor EIs, but now with high per capita research output) and join the universities occupying this quadrant from Figure 3."

> TD: *What appears to be unstated is that the entire set of 10 analyses reported are based upon four race categories alone, although there is bracketed comment: (ignoring gender imbalances). The consequence of this offhand remark is that the entire analysis appears to take* n=*4 rather than the claimed constitutional imperatives of race and gender, with* n=*8.*

This is an illuminating example of statements being taken out of context. The full statements regarding neglecting gender imbalances read as follows "Interestingly, if a university employs only black African female staff, for example, the EI is calculated to be a very poor 73.4. In the case of only black African staff (ignoring gender imbalances) this figure improves to 26.4, which is still far outside the acceptable tolerance levels."

Three paragraphs later we indicate: "In this study, we used national demographic data as indicated in Table 2. Thus the EI was calculated with respect to six categories (incorporating the four race groups and both genders." It is a pity that, when Dunne contacted us about our unpublished work (as he indicates in his Commentary), he did not see fit to clarify this matter. It would have saved him considerable unnecessary calculations. Fortunately Moultrie and Dorrington[3] were able to discern that we used n=6 in our study.

Concluding comments

In spite of the language used in some of the commentaries, the formula itself was not shown to be flawed. On the contrary, we were heartened to see that Moultrie and Dorrington[3] were able to use the formula effectively for their data.

We agree with all the commentators that the matter of transformation is a serious one that requires due consideration. In doing so, we urge readers to engage critically and objectively with our study. The basic parameters utilised must be understood and taken together with the conclusions reached, as with any scientific study.

Having made a case for a proper, detailed, reading of our original work, we must emphasise that we understand that our study (and the data it uses) can be nuanced. In particular, Moultrie and Dorrington[3] make some interesting points about how the data could be structured in order to reach more meaningful conclusions (although we commented earlier on some of the challenges with their suggestions). Indeed, in the presentations that we gave to various stakeholders and in some of the reviewers' comments, many of the issues they raise were also raised. It is clear that these issues are not settled and need further engagement until the full parameters under which the Equity Index can be employed are clear. Once those parameters are settled, the Equity Index can be utilised (together with other qualitative indicators) to provide a good picture of the success of transformation in South Africa.

We must comment, however, that the snapshot of the (lack of) success of transformation as evidenced by the EI was so stark that further nuances could not change the results significantly.

Given the snapshot of transformation that we have provided (or the modified version given by Moultrie and Dorrington[3]), it is clear that our government *and* universities' current transformation efforts are not going far enough or simply not being effective enough. Notwithstanding the many apparent criticisms levelled at our work, no one has been able to refute its fundamental (and rather obvious) conclusion: transformation is still a major challenge for the higher education sector in South Africa – it is 'painfully slow' and every institution has a challenge in this regard. This was clear from Soudien et al.'s[12] qualitative study in 2008. The EI results complement that work through the first quantitative indicator to aid transformation (and are in accordance with Soudien et al.'s[12] results). As we stated in our study '"Just as is the case with the CHET clusters, the system should engage with and decide upon important constituents (indicators) of the different indices". Indeed, this is in line with the recommendation[3] that "the focus of any metric of this sort should be on the progress of each institution within its circumstances". Our Equity Index can then be used, together with agreed benchmarks, to provide one quantitative means of monitoring the success of transformation.

Our purpose[5] was to show that the Equity Index previously introduced[4] could be used to quantitatively measure the success of transformation activities in an environment where no quantitative measure previously existed. We believe that we were successful in indicating that the Equity Index could be used in that way. In order to move this important matter forward, we do believe it would be best for the benchmarks and data to be further interrogated (by all concerned parties) so that the Equity Index could be used for the purpose it was intended – to help transformation in South Africa. We hope that the next publication on this subject is a peer-reviewed article using the Equity Index with agreed upon benchmarks and data which can guide the sector better.

Acknowledgement

We thank Prof. L. Nongxa (University of Witwatersrand) for his many useful comments that helped improve this Commentary.

References

1. Cloete N. A new look at demographic transformation: Comments on Govinder et al. (2013). S Afr J Sci. 2014;110(1/2), Art. #a0048, 4 pages. http://dx.doi.org/10.1590/ sajs.2014/a0048

2. Dunne T. Mathematical errors, smoke and mirrors in pursuit of an illusion: Comments on Govinder et al. (2013). S Afr J Sci. 2014;110(1/2), Art. #a0047, 6 pages. http://dx.doi. org/10.1590/sajs.2014/a0047

3. Moultrie TA, Dorrington RE. Flaws in the approach and application of the Equity Index: Comments on Govinder et al. (2013). S Afr J Sci. 2014;110(1/2), Art. #a0049, 5 pages. http://dx.doi.org/10.1590/sajs.2014/a0049

4. Govinder KS, Makgoba MW. An Equity Index for South Africa. S Afr J Sci. 2013;109(5/6), Art. #a0020, 2 pages. http://dx.doi.org/10.1590/sajs.2013/a0020

5. 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

6. Statistics South Africa. The 2011 census data [document on the Internet]. c2012 [cited 2013 May 28]. Available from: https://www.statssa.gov.za/Publications/P03014/P030142011.pdf
7. Department of Higher Education and Training. HEMIS database [accessed 2013 May 28].
8. Department of Higher Education and Training. 2011 Weighted research output. Pretoria: Department of Higher Education and Training; 2012.
9. Dorrington R, Bourne D, Bradshaw D, Laubscher R, Timaeus IM. The impact of HIV/AIDS on adult mortality in South Africa. A technical report. Cape Town: Medical Research Council; 2001. Available from: www.mrc.ac.za/bod
10. Ruegg W. The traditions of the university in the face of the demands of the twenty-first century. Minerva. 1992;30(2):189–205.
11. Stokes DE. Stating the problem. In: Pasteur's Quadrant: Basic Science and Technological Innovation. Washington DC: Publisher Brookings Institution Press; 1997. p. 1–25.
12. Soudien C, Michaels W, Mthembi-Mahanyele S, Nkomo M, Nyanda G, Nyoka N, et al. Report of the Ministerial Committee on Transformation and Social Cohesion and the Elimination of Discrimination in Public Higher Education Institutions. Pretoria: Department of Education; 2008.

The contribution of higher education institutions to the South African economy

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

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

CORRESPONDENCE TO:
Roula Inglesi-Lotz

EMAIL:
roula.inglesi-lotz@up.ac.za

POSTAL ADDRESS:
Department of Economics, University of Pretoria, Pretoria 0002, South Africa

KEYWORDS:
South Africa; higher education; universities; economic contribution

HOW TO CITE:
Pouris A, Inglesi-Lotz R. The contribution of higher education institutions to the South African economy. S Afr J Sci. 2014;110(3/4), Art. #a0059, 5 pages. http://dx.doi.org/10.1590/sajs.2014/a0059

We present the direct and indirect contributions of higher education institutions in South Africa to certain macroeconomic indicators such as GDP and employment, with the ultimate purpose of establishing their importance for the country. Taking this a step further, funding of these institutions is crucial in order for them to continuously produce outcomes in terms of research and skilled graduates. Hence, we compare the South African research and development (R&D) expenditure with international best practice. Policy implications are also discussed, especially in the light of the new funding formula for universities to be announced by the Department of Higher Education and Training.

Introduction

Tertiary education contributes to social and economic development through four major missions: the formation of human capital, the building of knowledge bases (primarily through research and knowledge development), the dissemination and use of knowledge (primarily through interactions with knowledge users) and the maintenance of knowledge (inter-generational storage and transmission of knowledge).

Based on robust evidence that human capital is a key determinant of economic growth and on emerging evidence that higher education is also associated with a wide range of non-economic benefits such as better health and well-being, governments internationally support the sector financially. Investment in human capital and, by implication, higher education has moved to the centre stage of strategies to promote economic prosperity, fuller employment and social cohesion during the last decade).[1]

Universities have historically always been of importance in the domain of knowledge. However, since the rise of mass education after World War II, higher education has changed the character of universities and they have become crucial for employment, social mobility, economic growth and economic development. Today the importance of a vibrant higher education sector is recognised internationally.

The 1980 report[2], 'Technical Change and Economic Policy', is now widely recognised as the first major policy document to challenge the macroeconomic interpretations of the 1970's crisis and to emphasise the role of technology in finding solutions. For example, innovation can be more powerful than wage competitiveness in stimulating an economy[3] and universities play a crucial role in the process.

A substantial volume of literature on the topic indicates that the private returns of higher education institutions to R&D are strongly positive and higher than those for capital and that the social returns are even higher.[4] KPMG[5] reported that 'increasing university funding from its current level of 1.6% of GDP to 2% of GDP in Australia and increasing the share of Commonwealth government grants up to 50% from 42%, led to a 5.8% gain in real GDP and a 5.2% gain in living standards in the long term'. Similarly, Universities UK[6] showed that universities contributed 2.3% of the UK GDP in 2008 and 'that the effectiveness of the higher education sector in generating impact is relatively high compared to other sectors of the economy'.

The 2010 edition of *Education at a Glance*[7] shows that public resources invested in education ultimately receive returns in even greater tax revenues. On average, across Organisation for Economic Co-operation and Development (OECD) member countries, a person with a tertiary level of education will generate an additional USD119 000 in income taxes and social contributions over their working life compared to someone with only an upper secondary level of education. Even after subtracting the public revenue that has financed the degree, an average of USD86 000 remains – almost three times the amount of public investment per student in tertiary education. The returns to society are even larger because many benefits of education (e.g. in terms of health) are not directly reflected in tax income.

The importance of higher education has never been more evident than in the recent (2008) international recession and financial crisis. Countries set R&D expenditure targets and use R&D expenditure as a stimulus for economic recovery. For example, the European Union has urged member countries to increase investment in R&D and consider ways to increase private sector R&D investments. The US government, as part of the *American Reinvestment and Recovery Act of 2009*, has increased its spending on R&D related to climate change by USD26.1 billion and to energy by USD6.36 billion. In addition, USD10 billion was allocated to the US National Institutes of Health (NIH) for biomedical research and an additional USD2.3 billion was allocated for research funded by the National Science Foundation.[8] The OECD[3] stated: 'Despite the slowdown in economic growth and the resulting fall in tax revenue government investments in R&D have outpaced outlays in other areas. Government investments or spending and tax cuts, taken together, have represented on average more than 3% of GDP in the OECD area and up to 5% of GDP in the United States and Korea'.

In addition, looking at the impact of academic research output to the economic growth, a number of authors[9-15] have concluded that there is a certain level of causality, but they do not come to an overall agreement and conclusion with regard to the direction and size of the influence. Lee et al.[11] argue that the direction of the causality depends highly on the developmental stage of a country: a weaker or no relationship was found for the developed economies in their study and a stronger relationship was found for the developing countries. The unambiguity in the direction of causality found in the international literature can be attributed to a country's level of economic growth and development or different periods examined or dissimilar academic and research systems. In South Africa, after a long period of decline and consolidation in the number of publications[16], appropriate incentives[17] have raised the universities' outputs[18].

In this analysis, we discuss direct and indirect contributions of higher education institutions in South Africa to certain macroeconomic indicators such as GDP and employment. After establishing the importance of the higher education sector for the economy, we present interesting facts on the R&D expenditure to higher education in South Africa and then position the country internationally. Finally, we derive policy implications of the analysis and provide a discussion on the future of higher education institutions in South Africa.

Contribution of higher education institutions to the economy

The higher education institutions in the country produce, in addition to knowledge and skilled graduates, their own output and employ numerous employees in different professions and at various qualification and skills levels. Universities also generate additional output and employment in other economic sectors through secondary or 'knock-on' multiplier effects. These effects comprise two types of economic interaction:

1. Indirect effects – universities purchase goods and services from other sectors in order to support their own activity, thereby stimulating activity within those industries. The supplying industries also buy from other suppliers in order to fulfill university orders, and those suppliers in turn buy from others, so there is a ripple effect.
2. Induced effects – universities pay wages and salaries to employees, who in turn spend this income on consumer goods and services. This spending creates wage income for employees in other sectors, who also spend their income and so on, creating a ripple effect throughout the economy as a whole.

There are two approaches that can be taken in order to produce estimates of these 'knock-on' effects. Either an operational model of the national economy can be developed and used, or the gross output multipliers can be estimated from international estimates. The sectoral gross output multiplier is the ratio of total output to direct output. The gross output multiplier for UK universities is estimated to be 2.38 and we use it for our estimates.[6]

Table 1 shows the impact of the higher education sector (23 universities) on the country's economy. The total impact as a percentage of the GDP in current prices amounted to 2.1% in 2009.[19] The release by StatsSA identified that cash receipts from operating activities of higher education institutions amounted to R36 892 million during 2009.[19] Following the definition, this amount is approximately equivalent to the higher education sector's gross output. From this figure, we estimate the sector's value addition and then add the multiplier effects.

Table 1: Impact of the higher education sector on South Africa's economy in 2009

	Universities (Rand billions)
Output	
Direct output	R36 892
Secondary output	R50 910
Total output generated (direct+secondary)	R87 803
GDP	
Direct GDP	R23 350
Secondary GDP	R25 455
Total GDP generated (direct+secondary)	R50 805

It should be mentioned that additional economic impact is generated as foreign students spend resources in the country (outside of universities) which would not have been spent in the absence of universities. Similarly, expenditure generated outside the sector from international conferences, visiting academic staff, etc. is not included in the estimation.

The importance of the above StatsSA[19] figures becomes apparent when they are set in the national context. Figure 1 shows that the value added by the higher education sector is just less than the contribution of the gold industry and substantially higher than the contribution of forestry, textiles, clothing and leather products, hotels and restaurants, and others.

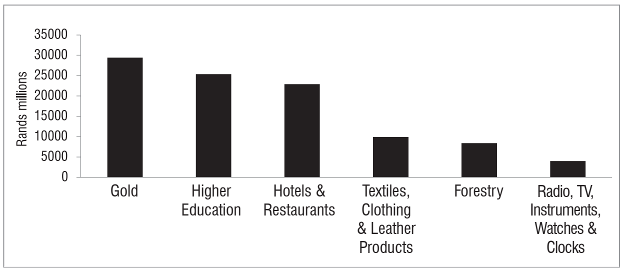


Figure 1: Value-added selected sectors for South Africa in 2009 (current values).

Table 2 shows that the university sector is a relatively high value added sector per employee. The sector's value added per employee is higher than that of the construction and agriculture sectors and just less than that of the manufacturing sector.

Table 2: Value added per employee

Industry	Value added (Rand millions)	Employees	Value added/ employees
Agriculture	R63 888	624 000	R102 384
Construction	R87 116	415 000	R209 918
Electricity, gas and water	R60 280	56 000	R1 076 428
Manufacturing	R330 310	1 185 000	R278 742
Transport, storage and communication	R199 065	359 000	R554 498
Universities	R25 350	113 000	R224 336

Source: StatsSA[19,22-23]

We have already mentioned that higher education makes an additional contribution to the economy through employment. In 2009, 112 797 staff (41 428 permanent and 71 369 temporary) were employed in public higher education institutions. This figure is up from 101 186 during 2004 and 108 697 employed during 2007. It should be emphasised that during the period 2008–2009 the rest of the economy lost 870 000 jobs.[20]

Estimating the total employment impact of the sector (direct and indirect), we identified that, during 2009, the sector employed 228 978 employees. This figure is substantially higher than the number of jobs available in the utilities sector (98 000) and slightly lower than the number of jobs in the mining and quarrying sector (296 000 jobs).[20]

Funding of higher education institutions in South Africa

In 2010, a document entitled the 'New Growth Path'[21] was published in which the importance of higher education was recognised. It was suggested in this document that a main strategy to achieve the country's objectives is 'Greater support for R&D and tertiary education linked to growth potential and developing South Africa as the higher education hub for the continent'[21].

It is stated in the New Growth Path that:

> *Our technology policy has four main thrusts:*
>
> - *Achieving targets for increased R&D: In line with current targets, raising public and private spending on R&D from 0.93% in 2007/8 to 1.5% in 2014 and 2% in 2018;*
> - *Increasing the number of patents from 91 in 2008 to 200 in 2014;*
> - *Increasing the number of professionals and technicians from the current seven per 10 000 people to 11.*
> - *This will require costed and phased proposals from the relevant departments (DST, DHET, EDD and NT)* (p.23)

Figure 2 shows the gross domestic expenditure on R&D for 1999 (points) and 2009 (columns) as a percentage of GDP for a number of OECD and non-OECD countries. South Africa was among the countries that spent the least on R&D based on their GDP (in 2001 and 2008). It can also be observed that there was no substantial change in this expenditure during the period from 1999 to 2009. We can see that countries such as Korea and Japan that have the highest percentages of the total OECD R&D expenditures are also those that have the highest percentage of GDP spent on R&D.

The specific higher education expenditure on R&D for 1999 and 2009 is presented in Figure 3. South Africa is at the lowest end of the group that spent less than 0.4% of GDP on higher education R&D.

It can be observed from Figure 4 that internationally the business sector is responsible for the majority of R&D expenditure and South Africa is no exception. For the majority of the countries, the next higher contributor is the higher education sector, whereas in South Africa, the higher education sector and the government contribute approximately the same to R&D expenditure in the country.

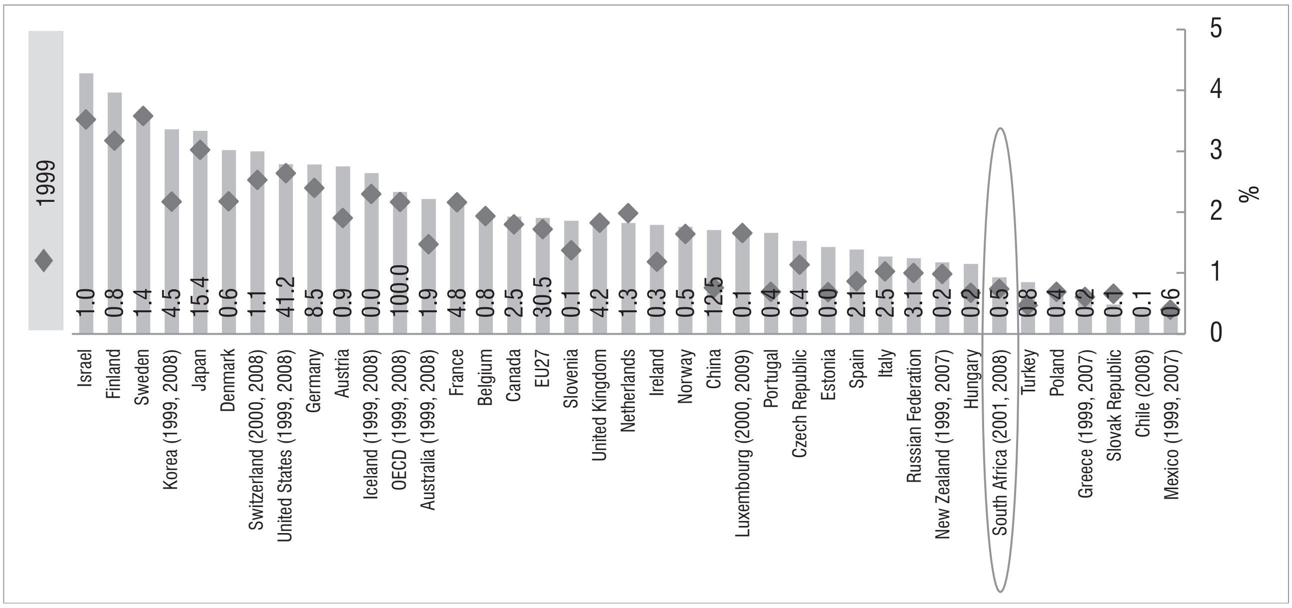
Policy discussion

Universities were identified as the main repository of knowledge in the country, producing more than 85% of the country's publishable research. Furthermore, it is estimated that the contribution of the higher education sector to the economy as it is manifested in the sector's added value is just less than the contribution of the gold industry and substantially higher than the contribution of the forestry, textiles, clothing and leather products, hotels and restaurants, and other sectors.

Estimating the direct and indirect economic impact of the country's 23 universities, it was identified that they contributed 2.1% of the country's GDP in 2009. In terms of added value, the contribution of the 23 universities is just less than the contribution of the gold industry and substantially higher than the contribution of the forestry, textiles, clothing and leather products, hotels and restaurants, and other sectors.

South Africa's 23 universities employed (directly and indirectly) 228 978 people in 2009. This figure is substantially higher than the number of jobs available in the utilities sector (98 000) and just less than the number of jobs in the mining and quarrying sector (296 000 jobs) in the country. Importantly, while the economy was shedding 870 000 jobs during the 2008–2009 period, the higher education sector showed a 3.8% increase in direct employment.

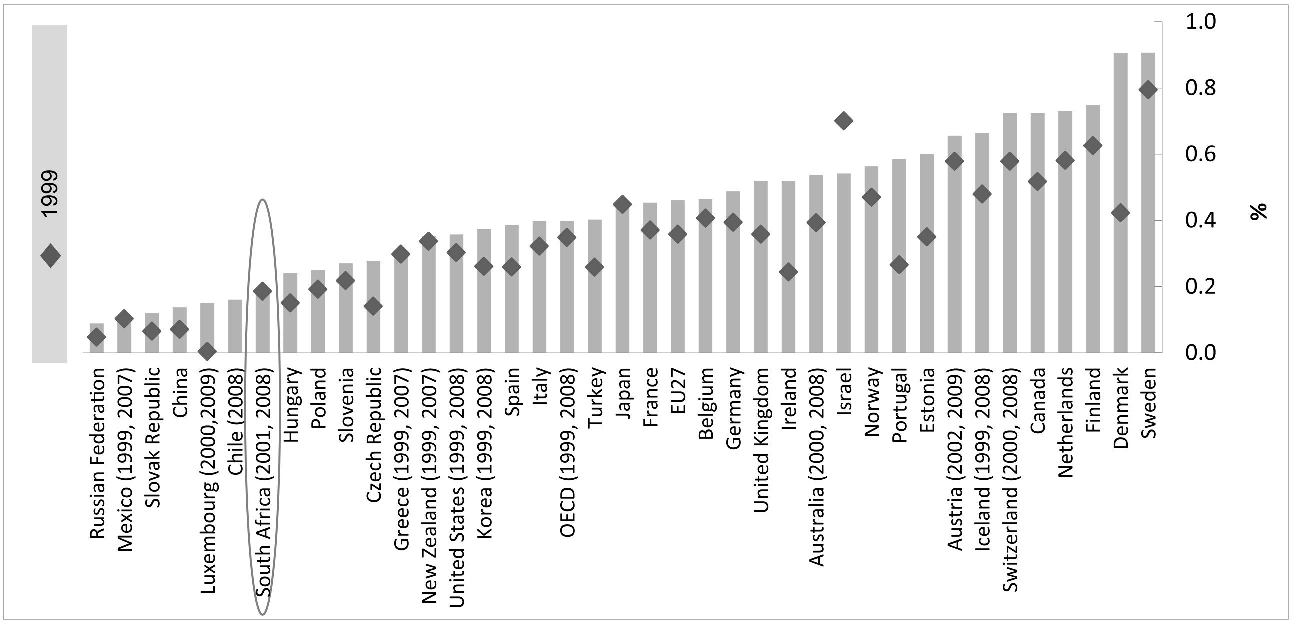
Universities are recognised internationally as cornerstones of knowledge-based societies and governments support them even during recessions or crisis periods. In the USA during the recent crisis, the government, as



Source: OECD[24]

Note: The figures at the bottom of the columns denote % of total OECD R&D expenditure, 2009; the rhombus symbols denote the 1999 values.

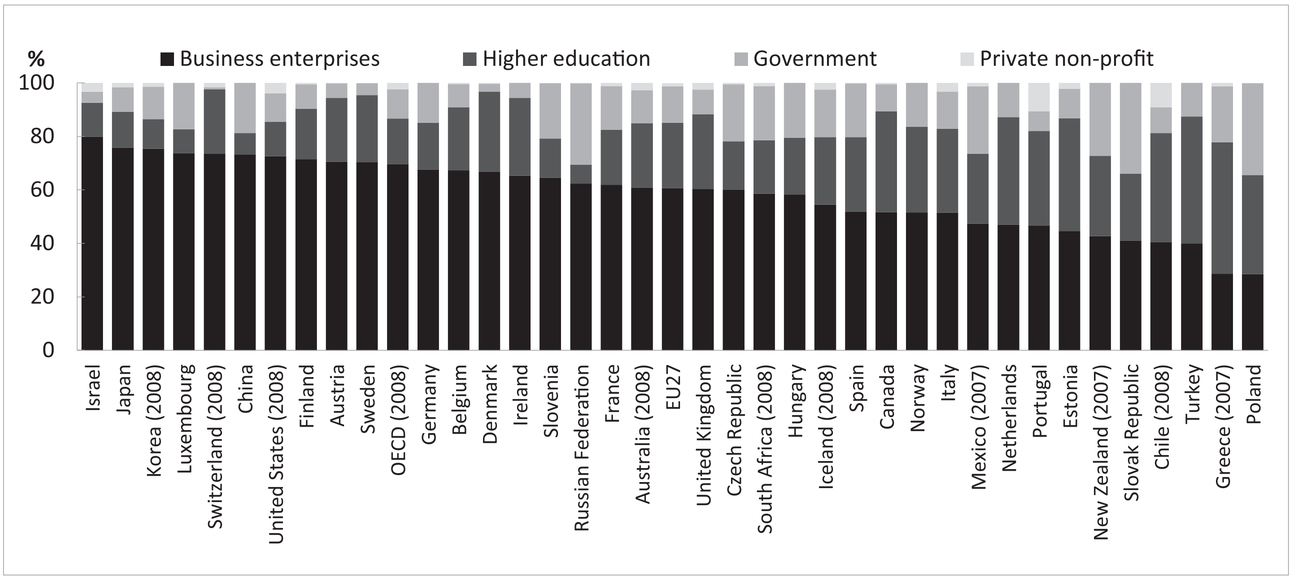
Figure 2: Gross domestic expenditure on R&D (% of GDP) in 1999 and 2009.



Source: OECD[24]

Note: the rhombus symbol denotes the 1999 values.

Figure 3: Higher education expenditure on R&D in 1999 and 2009.



Source: OECD[7]

Figure 4: R&D expenditure by performing sectors in 2009.

part of the A*merican Reinvestment and Recovery Act of 2009*, increased its spending on R&D related to climate change by USD26.1 billion and to energy by USD6.36 billion. In addition USD10 billion was allocated to the US NIH for biomedical research and an additional USD2.3 billion was allocated for research funded by the National Science Foundation.

Despite the multitude of socio-economic benefits offered by the sector and evidence from international best practice as manifested in the actions of OECD countries, the South African government is neglecting the higher education sector.

A comparison of South Africa with the OECD countries identified the following:

- South Africa's low R&D intensity, the relative low figure in targeting and the lack of relevant budgetary appropriations indicate that the science authorities in the country, despite the evidence from international best practice, need to accept in the future that without bold undertakings in the field of science and technology they undermine the country's socio-economic future.
- Government expenditure to universities in South Africa is well below the OECD average. If the country wishes to reach the OECD average the university allocations should be increased by at least 40% of the 2007 values, and if the country wishes to be among the top quartile of the OECD countries, the investment to universities should be doubled (from the 2007 levels).
- South Africa's higher education R&D intensity of 0.18% of GDP is well below half of the OECD intensity. This intensity should be increased by 100% if the country wishes to be comparable

with the average OECD country. This would require an additional investment of R3 billion per annum for R&D dedicated for the higher education sector.

- A total of 20.7% of R&D expenditure in the higher education sector arises from the local business sector. To put this figure in context, the country's science councils (Council for Scientific and Industrial Research, Medical Research Council, Agricultural Research Council, etc.) receive only 9.7% of their R&D expenditure from the business sector. Similarly, industry funds 6.2% of higher education's R&D activities in the OECD countries and 6.6% in the 27 countries of the European Union (EU-27).

The above findings raise a number of policy questions. For example, why does the government fund higher education institutions only to a limited extent although they produce 85% of the country's publicly available research? Is the current model of supporting R&D according to international best practice or it is the result of historical misalignment? What are the benefits from the current government support for private sector research?

The challenge for South Africa is clear. But so is the solution: evidence shows – consistently, and over time – that countries and continents that invest heavily in education and skills benefit economically and socially from that choice. For every rand invested in attaining high-skilled qualifications, taxpayers get even more money back through economic growth. Moreover, this investment provides tangible benefits to all of society – and not just to the individuals who benefit from the greater educational opportunities. The Department of Higher Education and Training is in the process of adjusting the funding formula for universities. It may be a unique opportunity for the country to recognise the benefits that the higher education institutions offer to the country and make appropriate decisions.

References

1. UNESCO. Financing education – Investments and returns: Analysis of the world education indicators. Montreal: UIS, OECD; 2003. Available from: http://www.uis.unesco.org/Library/Documents/wei02_en.pdf
2. Organisation for Economic Cooperation and Development (OECD). Technical change and economic policy: Science and technology in the new economic and social context. Paris: OECD; 1980.
3. Organisation for Economic Cooperation and Development (OECD). Science, technology and industry outlook 2010. Paris: OECD; 2010.
4. Hall HB, Mairesse J, Mohnen P. Measuring returns to R&D. NBER Working Papers Series 15622. Cambridge, MA: National Bureau of Economic Research; 2009.
5. KPMG. Economic modelling of improved funding and reform arrangements for universities. Canberra: Universities Australia; 2009.
6. Universities UK. The impact of universities in the UK economy. London: Universities UK; 2009.
7. Organisation for Economic Cooperation and Development (OECD). Education at a glance 2010. Paris: OECD; 2010.
8. Organisation for Economic Cooperation and Development (OECD). Policy responses to economic crisis: Investing in innovation for long-term growth. Paris: OECD; 2009.
9. King DA. The scientific impact of nations. What different countries get for their research spending. Nature. 2004;430:311–316. http://dx.doi.org/10.1038/430311a
10. Vinkler P. Correlation between the structure of scientific research, scientometric indicators and GDP in EU and non-EU countries. Scientometrics. 2008;74:237–254. http://dx.doi.org/10.1007/s11192-008-0215-z
11. Lee L-C, Lin P-H, Chuang Y-W, Lee Y-Y. Research output and economic productivity: A Granger causality test. Scientometrics. 2011;89:465–478. http://dx.doi.org/10.1007/s11192-011-0476-9
12. Inglesi-Lotz R, Balcilar M, Gupta R. Time varying causality between research output and economic growth in US. Scientometrics. Forthcoming 2013.
13. Inglesi-Lotz, R, Chang T, Gupta R. Causality between research output and economic growth in BRICS countries. Qual Quant. Forthcoming 2013. http://dx.doi.org/10.1007/s11135-013-9980-8
14. Inglesi-Lotz R, Pouris A. The influence of scientific research output of academics on economic growth in South Africa: an autoregressive distributed lag (ARDL) application. Scientometrics. 2013;95(1):129–139. http://dx.doi.org/10.1007/s11192-012-0817-3
15. Ntuli H, Inglesi-Lotz R, Chang T, Pouris A. Does research output cause economic growth or vice versa? Evidence from 34 OECD countries. J Assoc Inform Sci Technol. Forthcoming 2013.
16. Pouris A. The writing on the wall of South African science: A scientometric assessment. S Afr J Sci. 1996;92:267–271.
17. Pouris A. Effects of funding policies on research publications in South Africa. S Afr J Sci. 1991;87(3–4):78–81.
18. Pouris A. Science in South Africa: The dawn of a renaissance? S Afr J Sci. 2012;108(7–8):66–71.
19. StatsSA. Financial statistics of higher education institutions 2009. Statistical release P9103.1. Pretoria: StatsSA; 2010.
20. National Treasury. Budget review 2010. Pretoria: Treasury; 2010.
21. Presidency. The New Growth Path: The framework. Pretoria: Presidency; 2010. Available from: http://www.info.gov.za/view/DownloadFileAction?id=135748
22. StatsSA. Gross domestic product. Statistical release P0441. Pretoria: StatsSA; 2010.
23. StatsSA. Quarterly employment statistics. Statistical release P0227. Pretoria: StatsSA; 2010.
24. Organisation for Economic Cooperation and Development (OECD). Main science and technology indicators database. Paris: OECD; 2011.

BOOK 1 TITLE:
New South African review 3: The second phase – tragedy or farce?

AUTHORS:
Devan Pillay, John Daniel, Prishani Naidoo and Roger Southall

ISBN:
9781868147359

PUBLISHER:
Wits University Press, Johannesburg, R320

PUBLISHED:
2013

BOOK 2 TITLE:
State of the Nation. South Africa 2012–2013: Addressing inequality and poverty

EDITORS:
Udesh Pillay, Gerald Hagg, Francis Nyamnjoh, with Jonathan Jansen

ISBN:
978079692422

PUBLISHER:
HSRC Press, Cape Town, R295

PUBLISHED:
2013

REVIEW TITLE:
A rainbow, prophets and the vortex: The state of South Africa, *circa* 2012–2013

REVIEWER:
Iain Edwards

EMAIL:
edwards.researchsa@gmail.com

AFFILIATION:
Consulting historian

POSTAL ADDRESS:
PO Box 1145, Ruimsig 1732, South Africa

HOW TO CITE:
Edwards I. A rainbow, prophets and the vortex: The state of South Africa, *circa* 2012–2013. S Afr J Sci. 2014;110(3/4), Art. a0054, 5 pages. http://dx.doi.org/10.1590/sajs.2014/a0054

A rainbow, prophets and the vortex: The state of South Africa, *circa* 2012–2013

South Africa is now in the early months of important times.

In 2014 South Africans will celebrate four anniversaries. All are founding moments in this new democracy. The first two were in February. There was the 25th anniversary of State President F.W. de Klerk's 2 February 1990 State of the Nation address to the Joint Sitting of Parliament, announcing the unbanning of outlawed liberation movements and persons and the release of `political` prisoners. Then 11 February marked the 25th anniversary of Nelson Mandela leaving prison and delivering his landmark address from the balcony of the Cape Town City Hall. On 27 April 20 years ago, the 10th Parliament of the Republic was duly elected in South Africa's first universal general election. Finally, 10 May will mark the 20th anniversary of Mandela taking the oath of office as President of the Republic.

And yet, as these historic milestones draw closer, the lustrous sheen has faded on public narratives of heroic liberation struggle and the miraculous birth of the `Rainbow Nation`. Important new public discourses are developing. Influential voices seek a re-examination of South Africa's liberation past, its transition to democracy, and its post-apartheid condition. There are major scholarly and political critiques of the existing relationships among the ruling political party, the state, and South African citizens. All political parties, including the ANC in its 2012 centenary year of celebration, have acknowledged the need for new, innovative political thinking. And yet the ANC-led Congress Alliance is beset with inter- and intra-organisational politicking and feuding. From its liberal anti-apartheid and pro-democracy pasts, South Africa's leading party of opposition has now also established firm bases as a party of provincial and municipal government standing poised to make further electoral gains. A new oppositional political party has now formed from within the heart of black consciousness liberation traditions. There is considerable public discussion on the need for a new standard of morality in public life. South Africa's standards of leadership are decried, most particularly by revered figures with impeccable anti-apartheid and `Struggle` backgrounds. Within these very public critiques, current ruling elites across the spectrum, are described as self-interested, lacking in moral principle, ethical scruple and the required political vision.

This year South Africa will also hold its fifth general election. Much politicking will evoke nostalgic memories of that first optimistic moment, and evocative memory of the now recently deceased Mandela. The ANC will use these feel-good anniversary moments to maximum self-congratulating celebratory purposes. Whilst the ANC will be returned as the ruling party of national government, election results promise to reflect a substantial re-alignment of political forces – the first such major change since 1994.

So, in this year there will be many articles, books, conferences – national and international – documentaries, exposés, films, histories, memoirs, memorials, polemics, and much more. Central to all these anniversary-time offerings should be two appropriately posed questions: `When did apartheid end?` and `After 20 years, what's freedom's dividend?'

These two books are about South Africa's post-apartheid contemporary condition, ca. 2012–2013. The first has its roots in left-wing anti-apartheid discourses of the 1980s. The *New South African Review* – now in its third iteration – is the descendant of that influential earlier series, then published by Ravan Press – also a powerful anti-apartheid force in the 1980s. As then, the series' intellectual energies come largely from within the Sociology Department at the University of the Witwatersrand. The New series is published by Wits University Press. The second book – dramatically larger in both scale and intent – is from the Human Sciences Research Council (HSRC). The HSRC is a statutory body, established in 1968 as the national social science public policy research institute. This book is the sixth volume of their post-apartheid State of the Nation series.

The key to both these scholarly works lies in their respective sub-titles. The *New South African Review*'s is *`The second phase – tragedy or farce`*. Here is explicit reference to the `The Second Phase` – a key recent ANC political analysis promising more 'radical policies and decisive action' in implementing the national democratic revolution, as adopted at their conference in Mangaung in 2013.[1] This notion is now a central element in the ANC's 2014 Election Manifesto.[2] Significantly, the HSRC's theme is also unambiguously expressed in its telling sub-title: *`Addressing inequality and poverty`*. This phrase also comes from current ANC thinking – its `addressing poverty, unemployment and inequality` policy statement and slogan.[3] This mantra is now also a key aspect of the ANC's 2014 Election Manifesto.[2] So, both books nail their relevancies firmly to their respective masts.

Significantly, these books also have some similar lines of both enquiry and analytical perspective. They also share similar overall general outlooks. Both show a deep sense of concern that South Africa has now passed into another phase of its post-apartheid trajectory. But here the sense conveyed is very, very different to the confident `second phase` public exuberance of the ANC. It is these two book's overall senses of the current South African condition that lies at the centre of this review.

Both books are scholarly reviews, of a special type. Both are thematic and clearly the result of strategic thought. Their respective sub-titles are most revealing. Neither book is chronologically rigid. There is much of contemporary history and politics in both. This history concerns both apartheid and the pre-1990 history of liberation movements, pro-democracy struggles and post-1990 contexts: global, regional and national. All of this detail is vital. Yet both works are essentially studies in contemporary and applied social science. Significantly, despite these two works being, respectively, the third *New South African Review* and the sixth volume of the *State of the Nation*, both eschew narratives of progress. The books resonate with sombre concern, offering ominously forbidding perspectives.

In both works the key moment is the Marikana massacre. Indeed, although strangely it is not clearly identified, the *Review*'s cover image is of `The Man in the Green Blanket` – the 32-year-old Eastern Cape migrant and mineworker leader, Mgcineni Noki, who was one of the 34 shot dead on that fateful 16 August 2012.[4] And to be clear, in neither of these works is `Marikana` simply a metaphor. That killing ground is used as a high-definition prism through which to see vast swathes of historical and contemporary South African life – now a central moment in post-apartheid South Africa's very changed contemporary condition. Here, 'distilled to its essence' – to borrow Devan Pillay's phrase in the *Review* – is the crux of these books' findings.[5] Preferring the phraseology of liquid dynamics, Olive Shisana – the influential CEO of the HSRC – puts it clearly in her Foreword: South Africans 'have entered a `watershed`'[6]. In this scenario the central actor is the poor – those without a freedom dividend – who 'refuse to accept the persistence of inequality and poverty'[6].

The `Rainbow Nation` was a term defining a time of enormous optimism. Key to that South African optimism was trust in future governance, and faith in the human capacities and eager potentials of free people – South Africa's hugely and tragically diverse new citizenry in the making; its nascent civil society. To all these, Mandela's inaugural speech as President offered South Africans a social contract with their incoming government. His concluding remarks were[7]:

> *We understand it still that there is no easy road to freedom.*
>
> *We know it well that none of us acting alone can achieve success.*
>
> *We must therefore act together as a united people, for national reconciliation, for national building, for the birth of a new world.*
>
> *Let there be justice for all.*
>
> *Let there be peace for all.*
>
> *Let there be work, bread, water and salt for all.*
>
> *Let each know that for each the body, the mind and the soul have been freed to fulfil themselves.*
>
> *Never, never and never again shall it be that this beautiful land will again experience the oppression of one by another and suffer the indignity of being the skunk of the world.*

That optimistic sense of faith in human goodness and progress was quickly reflected in government policy. There was the Reconstruction and Development Programme. That same sense of optimism was also central to the new school education policy: outcomes-based education. And there was also Mandela's own pet project: the schoolchildren's feeding scheme. By the 2001 local government elections the entire country had been divided into metropolitan areas – the key home of the citizen. Indeed one can argue that the politico-legal death of apartheid came thus with these local government elections, with citizens now having to vote in these newly demarcated areas.

Broadly, the first was stymied by international financiers, with Mandela not using his undoubted moral authority to stand his ground; the second by its own ambition and a near complete lack of unionised teacher and civil service capacity, competence, and comprehension; and the third by greed – mainly of parents and schoolteachers. And the final one is now threatened by the popularly referred to `Traditional Courts` bill, pushed by none other than the current President. 'Let us solve African problems the African way, not the white man's way', Zuma recently told a meeting of traditional leaders.[8] The issue has been widely reported internationally.[9] To very many, here are apartheid's Bantustans being slipped in through the back door. Currently shelved – largely through opposition from ANC-led provinces – the ANC will re-introduce a revised form of this bill into Parliamentary processes after the forthcoming general election. Optimistic humane principles: non-racialism, non-sexism, and *ubuntu* – have become sterile mantras, devoid of real meaning. It is significant how little public emphasis ANC leaders give to the term `citizen`. Mandela mostly spoke, and most powerfully too, to 'My fellow South Africans'. Mbeki, in his sonorous addresses, preferred to stress white and black, and communities – advantaged and underdeveloped, rich and poor. Zuma's formal public face is opaque.

Forgetting downright antagonism and resentful meanness, within these optimistic times there were many words of wise counsel and considered caution. Considering the key ANC origins of the two sub-titles to these books, two cautions are vitally important. Significantly, one was never reported – never leaked – to the media. After the elections in 1994, at the ANC's first Parliamentary caucus, a member of that caucus remembers President Mandela (by then not a member of that caucus) cautioning as follows: 'It is only the ANC that can kill the ANC.'[10] Over time warnings continued to be expressed, even from the highest reaches within the ANC. One key public moment came in 2004. Addressing the Pan-African Parliament, President Mbeki made the following comments, the significance of which is impossible to miss. Mbeki[11] quoted Chinua Achebe:

> *They need nobody to remind them that they and their continent became an object of pity and despair among the peoples of the world, with their human dignity denied, because what happened and what we did suggest that we were incapable of doing the things we must do to restore and assert our own dignity, to do the things we must do to achieve the objective of a better life for all our people.*
>
> *Because they know what has happened to them, these masses understand very well why Chinua Achebe said the things he said in his classic work 'Things Fall Apart', when he wrote:*
>
> *'Warriors will fight scribes for the control of your institutions; wild bush will conquer your roads and pathways; your land will yield less and less while your offspring multiply; your houses will leak from the floods and your soil will crack from the drought; your sons will refuse to pick up the hoe and prefer to wander in the wilds; you shall learn ways of cheating and you will poison the kola nuts you serve your own friends. Yes, things will fall apart.'*

Mbeki's comments are an appropriate backdrop to Susan Booysen's excellent chapter on the ANC in the *Review*.[12] Her main interpretive lines are also used to good effect in Temba Masilela's important chapter in the HSRC volume.[13] Let's try and develop some of these points by asking suggestive comments and questions. We have to be brave.

It is clear that the ANC is now just a mature African nationalist party. It proclaims its permanence, and its prophetic messianic message, and godly derived authority. It draws heavily on an, albeit often invented or over-imagined, rich historical seam of legitimacy. It has very dynamic internal energies. Its membership swells. It dispenses patronage – within both itself and also into private and public sectors. It initiates, filters and limits policy discussions, and wider public discourses. Its street-fighting capacities remain. And by nature it is now a party of reform.

In the early 1990s, as the ANC went into complicated political and constitutional negotiations, it and its alliance partners were hugely divided over key approaches and beliefs. These divisions have not yet been resolved. Indeed, these differences were taken into government. The ANC has long debated the centres of power issue: mostly over whether there are one or two centres. Aren't there now many more centres of power? There is all too much of Tammany Hall in the ANC's senses of how entrepreneurial private enterprise joins political patronage, state tender and the civil service. Much patronage is actively planned and arranged on substantial scales. Much corruption operates in the public domain. Do these aspects of ANC organisational energy not count as organised criminal activity? Finally, regicide has been committed once – at Polokwane in 2007. Does one not acquire a taste for such excitement, plotting and deadly bloodletting?

These two books comprise 40 chapters, covering an enormous range of topics. There are clearly discernible themes common to both works. The first common theme concerns politics, power and governance. Here, importantly, the *Review* focuses heavily on the country's power elites. But there is an important difference. The HSRC volume avers such theoretical and political perspectives. Secondly, and again in both books, comes issues of the economy, labour, black economic empowerment (BEE), spatial developments and environmental issues. The third common theme covers public policy and social practice. It is here that the HSRC's long-standing depth and expertise in these areas comes into its own. In their work are extensive sociological perspectives, often highly detailed, of what it means to be living in South Africa at present. The final theme concerns South Africa and the wider world.

It is impossible to review all. There is a range of approaches and outlooks. For a variety of reasons there are stand-outs, needing mention. John Daniel's and Roger Southall's respective views on power elites are critically important.[14,15] William Atwell's work – a case study on BEE – is hugely stimulating, and encouraging too.[16] David Fig's perspectives on hydraulic fracking are refreshingly important.[17] So too is Mcebisi Ndletyana's view on politicking and policy incoherence.[18] Francis Nyamnjoh's article on seeing South African society through the prism of a Freedom Day broadcast from SAfm is a quite superb example of the sociological imagination at work.[19] There are more in similar vein.

But there are niggles, and more too.

These two works are both scholarly studies of very contemporary issues. Through all the chapters mention is made of a range of people: from elites to media commentators; political activists; trade unionists; foreigners; the sick; workers; the poor, homeless and destitute; victims of homophobic attack (gays and lesbians); xenophobes; abused women and children; shackland militants; organised criminals; and so forth. These are not just sociological categories. These are all living people. Both books have indexes. These are replete with individual names. Most of these named and indexed people live. Many are public figures. Many own, legislate or govern. Others do not. They all live and act. Yet in the *Review* only 1 chapter out of 16 – Fig's on fracking – really illustrates an understanding of the importance of oral interview source material and analysis. This too is the case in the HSRC's volume. Aside from a chapter with a single telephonic interview, and Njamnjoh's important tuning in to SAfm, only 1 chapter out of 24 makes use of oral interview material: the chapter by Laetitia Rispel and Jeanette Hunter on health-care reforms.[20] Surely this is inexplicable and a substantial methodological gap? What has happened to oral interviewing – of elites, and ordinary people too, in South African social science research methodology? What has happened to that dynamically rich and politically powerful tradition of oral interviewing – the insistence on the importance of the people's voice – that was so much a part of left-wing social scientist's politics and method during the 1980s? In both books, most of the cited works are secondary sources. We are not reading self-referential, disengaged but cleverly self-serving articles positioned for a proclaimed wider greater good, are we? Let's return to this issue in concluding remarks.

Review books like these two do present a range of chapters, mixing micro-studies and larger more comprehensive perspectives and themes. Readers tend towards the permissive, allowing for certain diversities and idiosyncrasies. But these are nevertheless reviews. They are not editions of a scholarly journal; nor are they published collections of conference papers.

Over the last few years there has been much discussion, in political, media and scholarly contexts, of political re-alignments. Much interest has focused on the growth of social movements and their potential importance within both local and national politics and governance. The *Review* has one chapter on social movements. This chapter concerns the Anti-Privatisation Forum in Johannesburg. Here are interesting personal reflections on weaknesses and failures; and an attempt in May 2012 to rejuvenate the movement. 'About fifteen men and women met on the stairs of Museum Africa one Sunday morning.'[21] Although strikingly reminiscent of earlier South African Trotskyite political traditions, we are not told of the Forum's political persuasions. This is an important pity. A major question concerns potential linkages between social movements and wider political forces.

In Gauteng during 2012–2013 the most important anti-privatisation issue was not a small fractious group agitating for free utility services and encouraging illegal electricity connections in Soweto. It was the sustained and widespread public opposition to the privatisation of Gauteng's public freeway system. These oppositional energies came from a political party – the Democratic Alliance (DA), the ANC-aligned trade union federation – the Congress of South African Trade Unions (COSATU) and largely social media-based social movements – such as the Opposition to Urban Tolling Alliance (OUTA) and Tolhek. Campaigns involved many tens of thousands of people, legal challenges, petitions and freeway go-slows. The latter organised by both COSATU and, separately, motorbike clubs. The issue has by no means disappeared. But it is noticeable how wider political differences came between parties having common cause in this single-issue campaign. And yet this single issue campaign was of huge potential benefit to a wider public. So, the behaviour of its citizen opponents is also revealing. The privatising of freeways can be quite properly viewed as an unjust law. When will South Africa's citizens – both its working- and middle-class commuter components – see South Africa's history of successful non-violent civil disobedience as inspiration for new senses of citizen power? Gandhi developed his ideas of *satyagraha* here. The Defiance Campaign – a campaign against unjust laws – would not have sustained itself without such principles. Here are vital issues that need coverage which are not covered.

To return now to Shisana's analytical overview in the HSRC volume: a 'watershed' moment. As the two works point out, much of the contemporary crises have been foreseen, and 'a turnaround approach has become inevitable'[6]. This issue is critically important and needs close scrutiny and development. Continuing her theme Shisana sets out 'multidimensional challenges…governance, politics, international relations, migration, poverty, unemployment, human settlements, the economy, and lastly, understanding the country's security environment'[6]. Similarly in their general introduction, Nyamnjoh and Gerald Hagg, together with Jansen, provide a number of even lengthier interrelated factors which are involved in any of the given key themes.[22] Similarly, in the *Review*'s Introduction by Pillay, there is a section on Marikana. Herein is one absolutely key paragraph. It is one sentence, nine printed lines long. Here listed are the dynamics involved in creating the Marikana massacre. Each listed point is a highly respected sociological field of enquiry. There are 29 in number.[5]

This general sense is common to both works. Yet it needs further definition – certainly so to make the seriousness of the challenges that much more starkly clear. Shisana speaks of watersheds. But South Africa's current condition is far, far more serious. Keeping with the language of fluid dynamics, here now is the vortex: an irresistibly engulfing, spiralling and whirling mass. A watershed moment has long passed. How this present moment changes for the better is unclear.

Government goes into 2014 with the National Development Plan 2030, a massively researched and detailed plan of intent and purpose, derived from substantially consultative processes, for the next 15 years.[23] Although formally approved by the ANC, it remains hugely contested within both the ANC and its alliance partners. With this year being an election year, with an incoming administration and promised changes to the structure of executive government, delays in implementation are assured. And, the very idea of a 15-year plan excites the aspirations of the politically ambitious and the newly noticed too in any majority party. They have issues to prove and interests to serve. There is little unusual in this.

It was these issues which prompted the HSRC scholars to reflect on wider issues. South Africa has emerged from its `age of innocence`. Mistakes have been made. There are important lessons to be learnt. The legacies of the apartheid period are deeply entrenched. Leaders, across the spectrum, idealised all and also often poor and inappropriate visions, and their own celebrity and perceived omnipotence to advise, design, plan and introduce (or compel or impose) and then monitor change. Very basic errors have been made.[22] So, obviously, South Africans should be seeking and demanding – needing – a very thorough and honest understanding of how this came to be? In this regard Shisana comments as follows[10]:

> *More than ever there is a need for informed and robust debate that will enable public intellectuals, policymakers, academics, business, labour and civil society, to engage with our current challenges in a manner conducive to finding solutions.*

And that's it? This revealing reflection is heavily biased towards leaders and wider elites – including planners, policy wonks, applied social scientists and scholars in general.[6] But more of the same really will not do. Isn't there any sense within social scientists of a need to evaluate their own role in this state of affairs? Surely perspectives must be broadened. To repeat social scientist's own admissions: basic errors have been made. So, before such – continuing – meetings and debate why not also make a start to listen and learn from South African citizens? Isn't this basic ethically rooted and methodologically rigorous disciplined social science enquiry, very smart politics, and good humane common sense too? Don't just meet with fellow elites. Conduct probing recorded social science interviews with elites; and so too with ordinary citizens. Surely the very vibrancy of bravely engaged independent social science enquiry can give so much to responsible public policy discussion and decision?

Despite its professed centrality, neither book has a chapter on Marikana. This is a glaring omission. The very soon before he was shot dead Noki features largely, although unnamed, on the *Review*'s front cover. He doesn't even get entry mention in the HSRC's index. So, there is an obvious and good place to start – just have the basic common decency and respect to acknowledge the name of the now dead `Man in the Green Blanket`.

One is reminded of Mbeki's reciting from Achebe; and then of Yeats's `The Second Coming`:

> *Turning and turning in the widening gyre*
> *The falcon cannot hear the falconer;*
> *Things fall apart; the centre cannot hold;*
> *Mere anarchy is loosed upon the world,*
> *The blood-dimmed tide is loosed, and everywhere*
> *The ceremony of innocence is drowned;*
> *The best lack all conviction, while the worst*
> *Are full of passionate intensity.*[24]

Throughout these two books scholars are expressing sentiments so very similar to ones offered, almost daily, in wider public discourses. This is despite these scholar's own findings raising warning bells. And, although some writers within these books prefer otherwise, it is wise to not easily dismiss such public sentiments. We are all, after all, only human. To paraphrase a prevalent public view, South Africans require an inspirational leader of moral legitimacy forging a rejuvenated political discourse around a core common vision. There is a huge degree of observational and emotional sense in such an outlook. But, sadly, isn't the dystopia of the vortex a lot more vigorously stronger and persistently dynamic and powerful than such utopian wishes? Do South Africans just need a `Lula moment`? [25] Or do South Africans really seek to create another man-god? Are South Africans not caught in a post-Mandela prophetic quest – seeking the new Messiah with a magic wand? And until then, just how do they intend awaiting the Second Coming?

References

1. African National Congress. 53rd National Conference Resolutions [document on the Internet]. c2013 [cited 2014 Mar 13]. Available from: http://www.anc.org.za/docs/res/2013/resolutions53r.pdf
2. Zuma J. Together we move South Africa forward: 2014 Election Manifesto [document on the Internet]. c2014 [cited 2014 Mar 13]. p.3. Available from: http://www.anc.org.za/2014/wp-content/themes/anc/downloads/Manifesto_Booklet.pdf
3. African National Congress. Statement of the National Executive Committee on the occasion of the 102nd anniversary of the ANC; 2014 Jan 08.
4. Saba A. Remembering Marikana: The man in the green blanket. City Press. 2012 Sep 08;1.
5. Pillay D. Introduction. In: Daniel J, Naidoo P, Pillay D, Southall R, editors. New South African review 3: The second phase – tragedy or farce? Johannesburg: Wits University Press; 2013. p. 1–2.
6. Shisana O. Foreword. In: Pillay U, Hagg G, Nyamnjoh F, Jansen J, editors. State of the nation. South Africa 2012–2013: Addressing inequality and poverty. Cape Town: HSRC Press; 2013. p. xv.
7. Mandela N. Statement by Nelson Mandela at his inauguration as President. 1994 May 10.
8. BBC News Africa. Zuma says South Africa needs traditional courts [homepage on the Internet]. c2012 Nov 2 [cited 2014 Mar 13]. Available from http://www.bbc.com/news/world-africa-20178992.
9. Powell A. S Africa's traditional courts bill criticized [homepage on the Internet]. c2012 Oct 12 [cited 2014 Mar 13]. Available from http://www.voanews.com/content/south-africas-traditional-courts-bill-criticized/1525304.html.
10. Confidential source. Interviewed by Edwards I. 2002 Feb 06.
11. Mbeki T. Address by the President of South Africa at the Pan-African Parliament; 2004 Sep 16; Gallagher Estate Midrand, Johannesburg, South Africa. Available from: www.dirco.gov.za
12. Booysen S. The ANC circa 2012–2013: Colossus in decline? In: Daniel J, Naidoo P, Pillay D, Southall R, editors. New South African review 3: The second phase – tragedy or farce? Johannesburg: Wits University Press; 2013. p. 39–60.
13. Masilela T. A review of the national development plan: Introducing politics and the state in South Africa. In: Pillay U, Hagg G, Nyamnjoh F, Jansen J, editors. State of the nation. South Africa 2012–2013: Addressing inequality and poverty. Cape Town: HSRC Press; 2013. p. 35–50.
14. Daniel J. Party, power and class. In: Daniel J, Naidoo P, Pillay D, Southall R, editors. New South African review 3. The second phase – tragedy or farce? Johannesburg: Wits University Press; 2013. p. 17–38.
15. Southall R. The power elite in democratic South Africa; race and class in a fractured society. In: Daniel J, Naidoo P, Pillay D, Southall R, editors. New South African review 3. The second phase – tragedy or farce? Johannesburg: Wits University Press; 2013. p. 61–75.
16. Atwell W. Broad-based BEE? HCI's empowerment model and the syndicalist tradition. In: Daniel J, Naidoo P, Pillay D, Southall R, editors. New South African review 3. The second phase – tragedy or farce? Johannesburg: Wits University Press; 2013. p. 138-153.

17. Fig D. Hydraulic fracturing in South Africa. Correcting the democratic deficits. In: Daniel J, Naidoo P, Pillay D, Southall R, editors. New South African review 3. The second phase – tragedy or farce? Johannesburg: Wits University Press; 2013. p. 173–194.

18. Ndletyana M. Policy incoherence: A function of ideological contestations? In: Pillay U, Hagg G, Nyamnjoh F, Jansen J, editors. State of the nation. South Africa 2012–2013: Addressing inequality and poverty. Cape Town: HSRC Press; 2013. p. 51–71.

19. Nyamnjoh F. Freedom Day on SAfm. Introducing South African society. In: Pillay U, Hagg G, Nyamnjoh F, Jansen J, editors. State of the nation. South Africa 2012-2013: Addressing inequality and poverty. Cape Town: HSRC Press; 2013. p. 305–332.

20. Rispel L, Hunter J. The Millenium Development Goals and healthcare reforms in South Africa: Progress and challenges. In: Pillay U, Hagg G, Nyamnjoh F, Jansen J, editors. State of the nation. South Africa 2012–2013: Addressing inequality and poverty. Cape Town: HSRC Press; 2013. p. 536–445.

21. Veriava A, Naidoo P. Predicaments of post-apartheid social movement politics: The Anti-Privatisation Forum in Johannesburg. In: Daniel J, Naidoo P, Pillay D, Southall R, editors. New South African review 3. The second phase – tragedy or farce? Johannesburg: Wits University Press; 2013. p.76–89.

22. Nyamnjoh F, Hagg G, Jansen J. General introduction. In: Pillay U, Hagg G, Nyamnjoh F, Jansen J, editors. State of the nation. South Africa 2012–2013: Addressing inequality and poverty. Cape Town: HSRC Press; 2013. p. 1–31.

23. National Planning Commission. National Development Plan Vision for 2030 [homepage on the Internet]. c2011 [cited 2014 Mar 13]. Available from: www.npconline.co.za

24. Yeats WB. The second coming (1919). In: Yeats WB. Michael Robartes and the Dancer. Shannon: Irish University Press; 1970 [reprint].

25. Munusammy R. Cosatu Congress: Vavi's time of reckoning and the elusive `Lula moment'. Daily Maverick. 2012 Sep 19.

BOOK TITLE:
Higher Education for the Public Good. Views from the South

EDITOR:
Brenda Leibowitz

ISBN:
9781920338886

PUBLISHER:
Trentham Books (Stoke on Trent, UK) in association with SUNMeDIA Stellenbosch (Stellenbosch, South Africa), R250

PUBLISHED:
2012

REVIEW TITLE:
Taking up the challenge of implementing higher education *for* the public good

REVIEWER:
Merridy Wilson-Strydom

EMAIL:
WilsonStrydomMG@ufs.ac.za

AFFILIATION:
Centre for Research on Higher Education and Development (CRHED), 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. Taking up the challenge of implementing higher education for the public good S Afr J Sci. 2014;110(3/4), Art. a0055, 2 pages. http://dx.doi.org/10.1590/sajs.2014/a0055

Taking up the challenge of implementing higher education *for* the public good

Is higher education for the public good still relevant in our global context where rankings, marketisation and an increasing focus on the instrumental value of higher education for economic growth often dominate at the levels of policy and practice? This timely book presents a good case for why it remains essential that all working in the higher education sector continue to interrogate the public good roles of our universities and our own practice. The purposes of higher education have been hotly debated for many years[1], and the extent to which we can and should view higher education as a public good is a central theme in these debates[2]. Brenda Leibowitz's edited collection is refreshing in its attempt to both theoretically and practically unpack the notion of what it means for higher education to work *for* the public good. In the introduction Leibowitz explains an important nuance of the title – a nuance that places the debates raised in the book somewhat outside of the contested distinction about the extent to which higher education is indeed a public good. Instead, this book is premised on the normative assumption that universities should work *for* the public good, and particularly so in the global South. The title seeks to

> *convey something more intentional and deliberate. It is the idea that we can conduct the three roles of higher education – research, teaching and community interaction – in such a way that we reflect on who higher education is for, who it can serve, and how.* (p.xxiii)

As such, the notion of the public good put forward in this book is 'associated with how people or groups behave' (p.xxii). In this case, people or groups refer to students and lecturers.

The book has its roots in a project run at Stellenbosch University, entitled 'Critical Professionalism' (Chapter 13) and as such, most of the authors hail from this university. Despite this apparent bias, a diverse selection of authors has contributed chapters, with perspectives from universities in the UK, Spain and the USA, as well as voices from three different South African universities. Several chapters draw on the work of Martha Nussbaum who has argued cogently for the importance of higher education – the humanities in particular – in the formation of democratic citizens.[3,4] However, the rich traditions of several important social theorists, from different disciplines, underpin the various chapters making the book theoretically engaging. In particular, the chapters included in this collection draw on the ideas of bell hooks, Amartya Sen, Martha Nussbaum, Nancy Fraser, Joan Tronto, Paulo Freire and Jurgen Habermas. Also important is the inclusion of research covering multiple disciplines, such as teacher education, social work, engineering, theology, visual arts, occupational therapy, psychology and mathematics education, as well as academic development. This multidisciplinary focus is a particular strength of the book as it provides a means for academics across disciplines to engage with the material presented and to reflect on what higher education for the public good might mean for their own practice.

The collection is presented in four parts. The first part (Chapters 1–4) tackles theoretical and policy-oriented arguments about higher education and the public good including the tensions between market responsiveness and social responsiveness; the second (Chapters 5–7) turns to the public good within an institutional context; the third (Chapters 8–12) considers teaching and learning or public good debates within the disciplines and the classroom; and the final section (Chapters 13–16) draws attention to the role of academics in fostering higher education for the public good. It is not possible to provide an overview of all 16 chapters in one review, although each has an important contribution to make to the whole. Instead, I would like to draw attention to a specific theme that runs across most chapters and which I think is of particular importance in the South African context – the role of teaching and curriculum in support of transformation and public good outcomes. It is to this theme that the remainder of this review turns.

In Chapter 4, Lange shows how the purpose of higher education 'to contribute to the socialisation of enlightened, responsible and constructively critical citizens' that is listed as one of the four purposes in the 1997 White Paper (and echoed in slightly different wording in the newly released 2013 White Paper for Post-School Education and Training[5]) has received relatively little attention to date. She notes further that evidence shows that universities in the country have not made sufficient progress in preparing students for a democratic society, within curricular and extracurricular spaces. This challenge is taken up by several contributors who demonstrate how curriculum and teaching and learning can be used for this purpose. Walker (Chapter 6) proposes an approach to curriculum that is grounded in the human development and capabilities approaches. Drawing on research focused on the role of professionals in reducing poverty, Walker outlines an approach to curriculum that would ensure that

> *students would then learn not only a curriculum of knowledge and skills, but the difference between simply having knowledge and skills on the one hand and, on the other, having the commitment and values to use these to the benefit of others.* (p.84–85)

A rich concrete example of the use of curriculum to encourage health sciences students to reflect on and understand their pasts and the implications thereof is shared by Nicholls and Rohleder in Chapter 9. These authors used a pedagogy of hope (also applied in Chapter 8 in the context of teacher education) to construct a curriculum that brought together final-year occupational therapy, psychology and social work students from Stellenbosch University and the University of the Western Cape. With their vastly different histories and current contexts the selection of these two universities was intentional in an effort to bring together students who might otherwise have little knowledge of each other's contexts and experiences. The programme leaders' concern that the students' future professional work would include interaction with clients from all racial groups and contextual backgrounds, yet their university experiences were not providing the skills needed, was the reason for constructing the course across these two different universities. The chapter presents in detail the curriculum as well as the pedagogies used and so provides a very useful case study for other academics wishing to embark on similar courses. It also demonstrates the power of participatory forms of pedagogy. Similarly, Boni et al. (Chapter 11) emphasise that *how* students learn can be more important than *what* they learn when seeking to foster cosmopolitan abilities and so also highlight the importance of participatory developmental pedagogies. Working on a global citizenship course offered to engineering students, these authors make use of the technique of moral dilemma and respectful dialogue as a means of requiring students to reflect on controversial topics within the field of development. On the basis of data collected from their students during an explorative study, Boni et al. conclude:

> [M]*any students showed clear indications of becoming empowered, being made aware of their capabilities and knowing they can influence what is happening, not only at the present time but also when they finish their studies and have an opportunity to put these cosmopolitan abilities into practice in their professional lives.* (p.148)

Despite inspiring examples of what is possible, the reader is also provided with a realistic sense of how difficult teaching for the public good can be. Particularly vivid is the chapter by Costandius (Chapter 15) in which she presents an extremely honest personal reflection on her 'embodied and emotional' (p.199) experiences of facilitating a course on citizenship in the visual arts. In my view, this realism and willingness to recognise the tensions and difficulties faced when teaching for the public good is an important contribution that this book makes, as common reasons given for not adopting such practices are, at least to some extent, pre-empted and addressed through their explicit recognition. As Nicolls and Rohleder remind us in Chapter 9:

> [T]*here is no quick fix, and everyone involved (educators and learners) is required to explore their identity markers (given and chosen) in an emotionally authentic manner. There can be many conscious reasons given for not doing this type of work (e.g. timetable issues, the high staff-student ratio required), but these reasons may hide a deeper unconscious denial of the painful past.* (p.125)

While recognising that it is not easy, this book calls on the university community, and particularly academics, to honestly and openly reflect on their roles in building a democratic society, through their work within universities. In support of those seeking to become critical professionals, Leibowitz and Holgate (Chapter 13) present seven aspects to take into account in building critical professionalism. In the Afterword, Leibowitz emphasises the need for safe spaces in which such work can take place because 'teaching for the public good can entail long and difficult work, and requires individuals to be prepared to experience moments of extreme vulnerability' (p.218). I strongly recommend this book and hope that more academics across disciplines will be inspired to embrace their vulnerability and take up the challenge of implementing higher education *for* the public good within their own contexts.

References

1. Kerr C. The uses of university. 5th ed. Cambridge, MA: Harvard University Press; 2001.
2. Longden B, Bélanger C. Universities: Public good or private profit. J High Educ Policy Manag. 2013;35(5):501–522. http://dx.doi.org/10.1080/1360080X.2013.825417
3. Nussbaum MC. Cultivating humanity. A classical defense of reform in liberal education. Cambridge, MA: Harvard University Press; 1997.
4. Nussbaum M. Not for profit. Why democracy needs the humanities. Princeton, NJ: Princeton University Press; 2010.
5. Department of Higher Education and Training. White Paper for Post-School Education and Training. Building an expanded, effective and integrated post-school system. Pretoria: Department of Higher Education and Training; 2013.

First complete account of the genus *Lachenalia* published

BOOK TITLE:
The genus *Lachenalia*

AUTHOR:
Graham Duncan

ISBN:
9781842463826

PUBLISHER:
Kew Publishing, Royal Botanical Gardens, Kew, GBP120 (hardcover)

PUBLISHED:
2012

REVIEW TITLE:
First complete account of the genus *Lachenalia* published

REVIEWER:
Brian W. van Wilgen

EMAIL:
bvanwilgen@sun.ac.za

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

POSTAL ADDRESS:
Department of Botany and Zoology, Stellenbosch University, Matieland 7602, South Africa

HOW TO CITE:
Van Wilgen BW. First complete account of the genus *Lachenalia* published. S Afr J Sci. 2014;110(3/4), Art. a0053, 1 page. http://dx.doi.org/10.1590/sajs.2014/a0053

South Africa is home to 6% of the world's approximately 370 000 plant species, making it the country with the richest temperate flora in the world. This dazzling diversity includes many large genera, and it is not often that a monograph appears that describes an entire, large genus. *Lachenalia* (also known as Cape hyacinths or *viooltjies*) is one such large genus. It has 133 known species that are confined to South Africa and (marginally) southern Namibia. These endemic plants have been popular with specialist bulb growers worldwide for over 100 years. The publication in 2012 of a comprehensive account of the genus marks the culmination of the life's work of two 20th-century South African plant taxonomists whose work between 1929 and 2012 has spanned more than eight decades.

Early records of *Lachenalia* date back to the late 17th century. In 1880, the Kew botanist John Baker published an account that described 27 species, divided among six genera. Baker later described more species, which culminated in 1897 in a monograph (published in the 6th volume of *Flora Capensis*) that recognised 42 species in five sub-genera. Most of the subsequent taxonomic work was done by Ms Winsome Barker, first curator of the Compton Herbarium at Kirstenbosch. Her first publication on the genus appeared in 1930, and over the next 59 years she described 47 new species and 11 new varieties. It was always her intention to publish a monograph on the genus, but the goal ultimately eluded her. Her last taxonomic description appeared in 1989, and she passed away in 1994 at the age of 87. In 1978, Graham Duncan took up a position at the Kirstenbosch Botanical Gardens, where he met and was influenced by Barker. Duncan is now Curator of the Bulbous Plants Living Collection at the Kirstenbosch Gardens, and over 35 years he has described a further 38 species of *Lachenalia*. In 1988, he published *The Lachenalia Handbook*, which illustrated and described 88 species. The handbook did not provide a comprehensive coverage of the genus, and was intended 'to collate available information so as to provide horticulturists and informed gardeners with a list of valid species names….and notes on identification and cultivation'. Almost another quarter of a century was to pass before the final goal of complete treatment was to be realised.

The information in this book arises from a combination of a great deal of searching in the field, horticultural efforts to grow and propagate specimens, and scientific endeavour. As a horticulturalist, Duncan has gained enormous insights from working with this genus for over three decades. He has combined this experience with scientific study, having recently completed a MSc degree that dealt with the cladistics of the genus, and which provided a sound basis for this book. Attention is also drawn to the work of others who have shown that differences in basic chromosome numbers result in breeding barriers between sympatric *Lachenalia* species whose flowering periods overlap. Thus, although many hybrids have been produced by horticulturalists, they are very rare in the wild. The book even contains a portrait of the 6-year-old Charles Darwin holding a potted hybrid *Lachenalia*, dated ca. 1816. The fact that certain species of *Lachenalia* exhibit a high degree of morphological variation has in the past led to confusion regarding their taxonomy, but the book points out that most species are in fact distinct and easily identified. Species with a similar (morphological) appearance are also not necessarily more closely related. In the 1988 *Lachenalia Handbook*, species were arranged by similar appearance (for ease of identification), and this has been interpreted by other scientists as indicating genetic relatedness. These questions are discussed in some detail, and provide a sound basis both for identifying species and understanding how they are related.

As is the case with so many Cape plants (the genus is concentrated in the southwestern Cape), many (over 50) are critically endangered, endangered or vulnerable, most of which is a result of habitat destruction for agriculture or urban development. Duncan recalls some of the 'highs' of his field career, including finding thousands of flowering specimens of *Lachenalia matthewsii* that had been considered extinct for over 40 years and finding an elusive specimen of the inconspicuous *L. maximilliani* under their parked car after a long and unsuccessful search of the surrounding area.

Although it has had a long gestation period, the resultant monograph has been worth the wait – it is a beautifully produced book. There are separate chapters on the history of the genus, cultivation and propagation, ecology and conservation, and biology. The largest section (over two-thirds of the book) is devoted to the taxonomic treatment; 11 new species are described for the first time. Each species is illustrated by means of full-colour photographs, and distribution maps are also provided. There is also a list of insufficiently known names (e.g. *Lachenalia cooperi* – 'Type not found, described from a cultivated plant'), and excluded taxa (previously described species that subsequently merged with other species). Other useful features include a key to the species, a table showing month-by-month flowering times for all species, and a glossary of terms for the uninitiated. The book is a blend of art and science, enhanced by 39 colour paintings of species by nine artists (17 of them by Barker). No taxonomic monograph is ever the final word – new species will be discovered, and changes to nomenclature will occur. However, *The Genus Lachenalia* will no doubt stand as one of the significant milestones of South African botanical publishing for decades to come. It will be a very welcome addition to the libraries of botanists, horticulturalists, conservationists and collectors of Africana.

Greenhouse gas emissions from shale gas and coal for electricity generation in South Africa

AUTHORS:
Brett Cohen[1,2]
Harald Winkler[1]

AFFILIATIONS:
[1]Energy Research Centre, University of Cape Town, Cape Town, South Africa
[2]The Green House, Cape Town, South Africa

CORRESPONDENCE TO:
Brett Cohen

EMAIL:
brett.cohen@uct.ac.za

POSTAL ADDRESS:
Energy Research Centre, University of Cape Town, Rondebosch 7700, South Africa

DATES:
Received: 24 Jun. 2013
Revised: 08 Oct. 2013
Accepted: 03 Nov. 2013

KEYWORDS:
shale gas; electricity; greenhouse gas emissions; South Africa; global warming potential

HOW TO CITE:
Cohen B, Winkler H. Greenhouse gas emissions from shale gas and coal for electricity generation in South Africa. S Afr J Sci. 2014;110(3/4), Art. #2013-0194, 5 pages. http://dx.doi.org/10.1590/sajs.2014/20130194

There is increased interest, both in South Africa and globally, in the use of shale gas for electricity and energy supply. The exploitation of shale gas is, however, not without controversy, because of the reported environmental impacts associated with its extraction. The focus of this article is on the greenhouse gas footprint of shale gas, which some literature suggests may be higher than what would have been expected as a consequence of the contribution of fugitive emissions during extraction, processing and transport. Based on some studies, it has been suggested that life-cycle emissions may be higher than those from coal-fired power. Here we review a number of studies and analyse the data to provide a view of the likely greenhouse gas emissions from producing electricity from shale gas, and compare these emissions to those of coal-fired power in South Africa. Consideration was given to critical assumptions that determine the relative performance of the two sources of feedstock for generating electricity – that is the global warming potential of methane and the extent of fugitive emissions. The present analysis suggests that a 100-year time horizon is appropriate in analysis related to climate change, over which period the relative contribution is lower than for shorter periods. The purpose is to limit temperature increase in the long term and the choice of metric should be appropriate. The analysis indicates that, regardless of the assumptions about fugitive emissions and the period over which global warming potential is assessed, shale gas has lower greenhouse gas emissions per MWh of electricity generated than coal. Depending on various factors, electricity from shale gas would have a specific emissions intensity between 0.3 tCO_2/MWh and 0.6 tCO_2/MWh, compared with about 1 tCO_2/MWh for coal-fired electricity in South Africa.

Introduction

South Africa is heavily dependent on coal for its primary energy supply, and is looking towards alternatives for electricity supply both in the interests of reducing the greenhouse gas (GHG) emissions intensity of its electricity supply and of seeking to diversify its energy sources. Shale gas potentially represents an alternative to coal for both electricity and energy supply. Reserves of shale gas in South Africa have been estimated to be the fifth largest globally, with current estimates placing reserves at 500 trillion cubic feet (tcf), although this figure can only be proved by exploration.[1,2]

When considering fuel combustion and electricity generation only, the GHG emissions associated with the production of electricity using gas (be it natural or shale gas) are definitely lower than those associated with electricity produced from coal. There have, however, been some studies that suggest that this may not be the case when considering the full life-cycle impacts, including those associated with fuel production and transport. In particular, fugitive emissions of methane increase the GHG impacts of this fuel option; fugitive emissions are those associated with leaks at various stages of fuel production such as during exploration, drilling and gas transport.

In this review, we summarise a selection of global studies to demonstrate some of the diverging findings on the GHG implications of shale gas, and highlight the critical assumptions that determine the results of such studies. The GHG emissions associated with the extraction of shale gas are then compared with those during electricity production from coal in South Africa.

The high contribution of fugitive emissions to the overall footprint of shale gas is in part because methane has a higher global warming potential (GWP) than carbon dioxide. In addition to exploring the magnitude of fugitive emissions, some studies have assessed the impact of assumptions around the time horizons of methane on the results, with assumptions of a shorter time horizon resulting in an even higher impact of fugitive emissions on the overall GHG footprint than when the GWP over a 100-year horizon is used. Some consideration is also given in this paper to the impacts of assumptions on GWP on the results.

It is recognised that both mining of coal and shale gas extraction have environmental and social impacts beyond those relating to their contribution to global warming, notably those on water resources. A discussion of such impacts is beyond the scope of this article.

Assumptions about global warming potentials and timescales

Prior to presenting the results from various studies on the GHG implications of shale gas versus coal, it is important to understand the science underlying one of the critical assumptions which has a substantial impact on the results: the GWP of methane, along with the timelines for which its impact is assessed.

Because multiple gases, including both carbon dioxide and methane, contribute to climate change, an emissions metric is needed to put emissions of different GHGs on a common basis when designing policies and measures to reduce GHG emissions. GWPs are one of the most commonly used 'conversion factors' for this purpose. The GWP is a multiplier that is used to convert a quantity of emissions of a gas species to an equivalent quantity of carbon dioxide (often referred to as CO_2-equivalent or CO_2e). Carbon dioxide is the major GHG in terms of overall contribution – despite its lower GWP relative to methane, the volumes emitted are much larger. This fact is most clearly seen in the contribution to radiative forcing, which for CO_2 is 1.66 W/m^2 (range 1.49–1.83 W/m^2) compared with a CH_4 radiative forcing of 0.48 W/m^2 (range 0.43–0.53 W/m^2).[3]

There are other very general formulations of an emissions metric (see for example Kandlikar[4]), but they suffer from difficulties, notably definition of the impact function and treatment of the time horizon. Other metrics, such as the Global Temperature Potential (GTP) have also been considered. GTP gives 'equivalent climate responses at a chosen time' and there is less focus on the fluctuations in the shorter term than when choosing a 20-year time horizon for methane.[5]

Hence the 'simpler and purely physical GWP index' was developed which compared the radiative forcing of a unit pulse, e.g. 1 kg, integrated over a time horizon (discussed further below). GWPs were also used in the Kyoto Protocol and its flexible mechanisms.[3] After comparing various options, the Intergovernmental Panel on Climate Change (IPCC) indicated that 'GWPs remain the recommended metric to compare future climate impacts of emissions of long-lived climate gases'[5]. The GWP is thus used here, and a discussion of other metrics is beyond the scope of this paper.

It is recognised, however, that several simplifications are made in developing the GWP multipliers. These simplifications, which are required in order to produce a workable emissions metric, have attracted criticisms of GWP. Among these are the impact function, but also (and most relevant to natural or shale gas) the assumptions about time horizons. The time horizon is particularly important in relation to gas, including shale gas. The residence time of CO_2, the most abundant GHG, in the atmosphere is between 50 and 200 years,[6] whereas methane remains for a much shorter period of 10 years.[7]

The IPCC's assessment reports have included GWPs for several gases and for different time horizons, with various factors leading to revision of these factors. The last comprehensive assessment of GWPs was undertaken by Working Group I of the IPCC in its Fourth Assessment Report (AR4). Working Group I focused on the physical science basis of climate change. The IPCC's AR4 reported the 100-year GWP of methane as 25 (the Second Assessment Report in 1995 suggested a lower value at 21) and the 20-year GWP of methane as 72 (56 in the Second Assessment Report).[6,8,9] The Fifth Assessment Report[10], due to be released in 2014, reviews these factors again. The studies cited here used GWPs from the Fourth Assessment Report. The Fifth Assessment Report provides updated GWPs (and other metrics), including GWPs over 20- and 100-year time horizons, with and without climate-carbon feedbacks (cc fb). For methane, AR5 reports GWP_{100} without cc fb as 28, with cc fb as 34, and GWP_{20} without cc fb as 84 and with cc fb as 86. Changing these values in our analysis would reduce their utility, as they would not be comparable to the literature; the general patterns reported hold.

The contribution of CO_2 and CH_4 to radiative forcing is well understood, with CO_2 making the greater contribution. The IPCC assessment[3] made it clear that the time horizon for assessment cannot be determined on a scientific basis alone, but that value judgements are required. In applying GWPs in a policy context, the crucial additional consideration in selecting a time horizon for the GWP metric is about the goal and purpose. Since that assessment report,[3] there has been political convergence on a long-term temperature stabilisation goal, within fairly narrow bounds of 2 °C, strengthening to 1.5 °C. Given that the goal of limiting temperature increase is now quantified, analysis can be aligned to the purpose of preventing the *long-term* increase in maximum temperature, which requires consideration of a longer time horizon and stabilisation of GHG concentrations (the objective of the United Nations Framework Convention on Climate Change[8]). Our view is that it seems advisable to retain the current practice of focusing on the GWP of non-CO_2 GHGs over a 100-year time horizon.

Life-cycle emissions from fuel supply

Sources of emissions from coal production

Coal production includes mining, preparation or beneficiation (depending on the coal quality and use) and coal transport. Coal mining is either conducted in underground or open-cast mines. GHG emissions sources from coal supply include:

- GHG emissions from electricity generation and diesel usage in mining. These emissions are primarily carbon dioxide although small volumes of other GHGs are also produced.
- Emissions associated with coal transport (electricity for conveyors, diesel for trucks and electricity and diesel for trains).
- Methane from coal seams that is released during exploration, venting of seams and ventilation of shafts.
- Spontaneous combustion of discards and disused workings, resulting in further CO_2 emissions.

While quantification of the first two sources is relatively straightforward, quantification of the methane emissions from coal seams and CO_2 emissions from spontaneous combustion is more difficult.

For underground coal mining, the majority of GHG emissions are in the form of methane that was previously retained in the coal seam. This methane is currently vented in South African mines. The remainder of the emissions from underground mines are GHG emissions associated with electricity use by the mine, and to a lesser extent combustion of liquid fuels. However, in open-cast mines, methane emissions released from the coal seam account for only a small portion of the total GHG emissions (less than 2%), with emissions associated with electricity use or liquid fuel use accounting for the majority of GHG emissions (depending on the mining method employed). De Wit[11] suggests that GHG emissions from coal mining and fugitive emissions are, however, small relative to those from combustion (6.55 Mt of CO_2-e emissions in 2003) and hence they are excluded from this current analysis.

Sources of emissions from shale gas supply

As indicated previously, emissions of methane can arise at various stages of the gas supply chain:

- well completion
- routine venting, equipment leaks and gas processing
- liquid unloading
- transport, storage and distribution

These sources of emissions are described in more detail below.

The World Resources Institute has produced a guideline for defining the boundary of which life-cycle impacts from shale gas production should be considered in determining the fuel impacts.[12] The World Resources Institute report reviews a number of studies to determine how extensively life-cycle impacts are incorporated into analyses. This review suggests that only one of 16 studies reviewed considered exploration, half considered site preparation and well completion, 11 considered fracking and 12 considered drilling. The substantial differences in system boundaries thus need to be taken into consideration when comparing different results, as illustrated further in this paper.

Emissions from well completion

During hydraulic fracturing ('fracking') to produce shale gas, fluid (mainly water) under high pressure is used to fracture deep shale formations. These fractures then begin releasing methane that was previously retained in the rock. The fracturing fluid is brought back to the surface over a period of between 5 and 12 days, after which methane flows freely from the well. Initially the amount of methane released during flowback is low, although it does increase significantly towards the end of the flowback period.

Different studies have produced diverging results on the amount of methane emitted during the flowback period. Howarth et al.[13] suggest that the gas emitted during this flowback period can be estimated from the well's initial production rate, and Jiang et al.[14] support this view, assuming that the rate of methane release during flowback is the same as that for the first 30 days of production. Cathles et al.[15], however, argue that the amount of methane produced during flowback is lower and cannot be estimated from initial production rates.

The majority of methane emitted during flowback is vented or flared – the latter clearly being preferred from a GHG emissions perspective as flaring converts methane to carbon dioxide, which has a significantly lower global warming potential. There are different assumptions on the proportion of methane emissions produced during flowback which are typically vented and flared. In their analysis, Howarth et al.[13] assumed that 85% of gas produced during flow back was vented and not flared, a conservative estimate in their opinion. Howarth et al.[13] support this high venting rate. Findings of other studies suggest that this venting rate is high. Both Jiang et al.[14] and Cathles et al.[15] report that 51–100% of the gas would typically be flared, and the remainder (0–49%) vented.

Emissions during extraction and processing

An extraction site has between 55 and 150 piping connections, all of which have the potential to develop gas leakages. Other equipment such as storage tanks and dehydration equipment have also been known to leak methane. In addition to unplanned leakages, certain pieces of equipment release gas to maintain internal pressure, and some pieces of equipment, such as pneumatic valves, release gas to power themselves when electricity is not available. Where gas requires further processing to increase purity, associated equipment also potentially releases emissions from leaks and during venting.

Emissions from liquid unloading

Liquid unloading is required when water and mud collect in the well and thus reduce the flow of gas. This process usually would occur later on in a well's life and would not be required for all wells. Collection of water and mud is a common problem in conventional natural gas wells but less of a problem in shale gas wells. During the unloading process, operators shut the well off to increase the pressure within the well. When the well is opened again, the pressure of the gas pushes liquid up and out of the well.[16] Gas is released during this process.

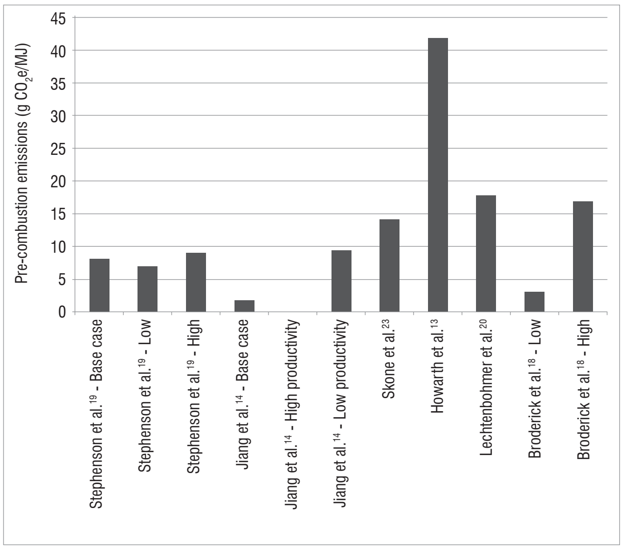
Emissions during transport, storage and distribution

Transmission, storage and distribution emissions are similar to those mentioned in the section on upstream extraction and processing, that is, those related to leaks and venting.

Quantifying the emissions from shale gas supply

A range of studies have been conducted which have focused on assessing the life-cycle emissions of shale gas, with the aim of determining the benefits of shale gas over conventional gas, coal and oil. Figure 1 shows a comparison among estimates of the emissions from shale gas supply up to the point of combustion, from a variety of different studies.[17]

From Figure 1 it can be seen that the majority of estimates of pre-combustion emissions are within a similar range, with the exception of the 'high' estimates of Broderick et al.[18], the estimates of Lechtenböhmer et al.[20] and those of Howarth et al.[13]. The Howarth et al. study received substantial attention as a result of its indications that emissions from shale gas were higher than those of other fossil fuels; however, a number of aspects of their analysis (including input assumptions and calculation errors) have come under criticism in subsequent studies, yielding significantly different results and questioning their higher overall GHG implications.[15,17] Howarth et al.[13] also chose to use a GWP for methane of 33, rather than that of 25 as used by other studies; however, this has been corrected for in Figure 1 to bring all of the studies on to the same basis with respect to GWP.



Note: *Jiang et al.[14] 'high productivity' and 'low productivity' refer to wells with high and low flow rates and lifetimes, respectively. The 'low', 'high' and 'base cases' in Stephenson et al.[19] refer to the extent of recovery which were explored by varying the impact of a number of different variables including ultimate recovery from the well, well completion emissions, flowback water, fractures per well and wellhead pressure.*

Figure 1: Pre-combustion emission estimates from shale gas (g CO_2e/MJ) obtained from different studies.

In interpreting the results presented in Figure 1, it needs to be recognised that the different studies do not all incorporate all of the steps in the pre-combustion supply of shale gas. Table 1 presents a breakdown of the steps included and excluded from the various studies. It also shows the substantially higher assumptions about well completion emissions included in the Howarth et al.[13] and Broderick et al.[18] studies. Note that the figures shown in Table 1 were calculated using a 100-year GWP for methane of 25, as per IPCC AR4.

Table 1: Breakdown of pre-production and pre-combustion emissions (g CO_2e/MJ) from shale gas

	Stephenson et al.[19]			Jiang et al.[14]			Skone et al.[23]	Howarth et al.[13]	Lechtenböhmer et al.[20]	Broderick et al.[18]	
	Base case	Low	High	Base case	High productivity	Low productivity				Low	High
Well completion	1.8	0.9	2.9	1.8	0.1	9.2	7.8	23.4	9	3	16.9
Venting, equipment leaks and gas processing	4.2	4.2	4.2				3.5	2.2	8.9		
Liquid unloading											
Transport, storage and distribution	1.9	1.9	1.9				2.7	16.2			
Pre-combustion total	7.9	7	9	1.8	0.1	9.2	14	41.8	17.9	3	16.9

Source: adapted from Forster and Perks[17]

Comparing electricity from shale gas and coal in South Africa

We have presented the outcomes of various other studies on the GHG emissions from shale gas. What is of particular interest in South Africa is comparing the range of potential emissions from using shale gas for electricity generation with emissions from coal-fired power generation, as coal is currently the predominant fuel for electricity generation, and there are emerging interests in shale gas in the country.

In 2012, Eskom produced 218 212 GWh from their coal-fired power stations (including own use), with associated emissions of 231.9 Mt CO_2, suggesting an emissions intensity of 1.06 tCO_2/MWh for coal-fired power alone.[21] The overall emissions intensity associated with electricity in South Africa is somewhat lower, because of the presence of other lower emissions options in the generation mix. It is the coal-fired portion only that is of interest in this current paper. As stated above, emissions associated with coal supply were considered negligible in this study.

Gas turbines, particularly combined cycle gas turbines, that are used to burn shale gas have higher efficiencies than coal-fired power stations. Combined cycle gas turbines in the USA have an average efficiency of 46%. Newer combined cycle gas fired power stations could achieve 55% or even 65% efficiency.[13,17,22] In the analysis here, the total life-cycle GHG emissions from producing a MWh of electricity are calculated by adding the pre-combustion emissions suggested by the various studies (as shown in Figure 1) to the emissions from burning the gas in a turbine. Consideration is given to the use of an older (46% efficiency), newer (55% efficiency) and potential future high efficiency (65% efficiency) power station.

Table 2 shows the life-cycle emissions intensity of electricity generated from shale gas calculated using this approach. Once again, it needs to be recognised that not all of the studies include venting, equipment leaks and gas processing, liquid unloading and transport, storage and distribution, and hence are not directly comparable but rather provide a range of results.

Table 2 shows clearly that under a wide range of assumptions about the efficiency of conversion of gas to electricity, and under all assumptions about the pre-production emissions of shale gas, shale gas has substantially lower GHG emissions than the 1.06 tCO_2/MWh associated with electricity produced from coal in South Africa. The lowest value in Table 2 is 0.31 tCO_2/MWh and the highest is 0.59 tCO_2/MWh. The differences relate to assumptions across studies of productivity, and the assumed efficiency of combined cycle gas turbines. In other words, how much lower the carbon footprint of shale gas is depends on assumptions about efficiency, fugitive emissions and GWPs. What is further interesting to note is the percentage of the emissions shown in Table 2 that are made up by pre-combustion emissions, and those that arise during electricity generation, using the assumptions about pre-combustion emissions in the various studies. These range from 1% in the Jiang et al.[14] base case, to 8% in the Skone et al.[23] case, to 34% in the case of Howarth et al.[13]

Finally, given the debates surrounding GWP for methane, and particularly the time horizon to be used, the calculations presented in Table 2 are repeated using a 20-year GWP for methane of 72.5 These results are shown in Table 3.

Clearly, the biggest impact of the shorter GWP is on the results of Howarth et al.[13], who assume the highest methane emissions associated with pre-combustion processing. However, it is interesting to see that even using a 20-year GWP for methane, shale gas has lower GHG emissions than coal-fired power in South Africa.

Conclusions

We have highlighted some of the life-cycle GHG considerations when comparing electricity generation from coal with that from shale gas in South Africa, and provided a comparison of emissions between the two electricity generation options. While a wide range of data is presented in the literature on the fugitive emissions from shale gas production, with not all data sources providing comparable information, the overriding indication is that shale gas has lower GHG emissions than coal-fired

Table 2: Calculated CO_2e/MWh electricity produced under different assumptions of combined cycle gas turbine efficiency using a 100-year global warming potential for methane (own analysis)

	Stephenson et al.[19]			Jiang et al.[14]			Skone et al.[23]	Howarth et al.[13]	Lechtenböhmer et al.[20]	Broderick et al.[18]
Assumed efficiency of conversion of gas to electricity	**Base case**	**Low**	**High**	**Base case**	**High productivity**	**Low productivity**				**Low**
46%	0.47	0.47	0.48	0.45	0.44	0.46	0.48	0.59	0.51	0.45
55%	0.40	0.40	0.40	0.38	0.37	0.39	0.40	0.52	0.43	0.38
65%	0.34	0.34	0.35	0.32	0.31	0.33	0.35	0.46	0.38	0.32

Table 3: Calculated CO_2e/MWh electricity produced under different assumptions of combined cycle gas turbine efficiency using a 20-year global warming potential for methane (own analysis)

	Stephenson et al.[19]			Jiang et al.[14]			Skone et al.[23]	Howarth et al.[13]	Lechtenböhmer et al.[20]	Broderick et al.[18]
Assumed efficiency of conversion of gas to electricity	**Base case**	**Low**	**High**	**Base case**	**High productivity**	**Low productivity**				**Low**
46%	0.53	0.52	0.54	0.46	0.44	0.50	0.54	0.88	0.63	0.47
55%	0.45	0.44	0.46	0.39	0.37	0.43	0.47	0.80	0.56	0.40
65%	0.40	0.39	0.41	0.33	0.31	0.37	0.41	0.75	0.50	0.34

power. This is the case even when considering a 20-year GWP time horizon for methane, which is not typical of the literature nor considered appropriate for the purpose of limiting long-term temperature increase. Depending on various parameters examined in this article, electricity from shale gas would have a specific emissions intensity between 0.31 tCO_2/MWh and 0.59 tCO_2/MWh, compared with about 1.06 tCO_2/MWh for coal-fired electricity in South Africa.

Authors' contributions

Both authors conceptualised the paper. B.C. was the principle author and was responsible for the quantitative assessments. H.W. authored the section 'Assumptions about global warming potentials and timescales' and provided editorial input on the remainder of the text.

References

1. Botha M, Yelland C. On fracking in the Karoo, open forums and the power of public opinion. Pretoria: EE Publishers; 2011.
2. De Wit M. The great shale debate in the Karoo. S Afr J Sci. 2011;107(7), Art. #791, 9 pages. http://dx.doi.org/10.1016/S0140-9883(96)00021-7
3. Intergovernmental Panel on Climate Change (IPCC). Climate change 2007: The physical science basis. Technical summary. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Geneva: IPCC; 2007. Available from: http://www.ipcc.ch/SPM2feb07.pdf.
4. Kandlikar M. Indices for comparing greenhouse gas emissions: Integrating science and economics. Energ Econ. 1996;18:265–282. http://dx.doi.org/10.1016/S0140-9883(96)00021-7
5. Intergovernmental Panel on Climate Change (IPCC). Climate change 2007: The physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Geneva: IPCC; 2007. Available from: http://www.ipcc.ch.
6. Intergovernmental Panel on Climate Change (IPCC). IPCC second assessment report. Geneva: IPCC; 1995. Available from: http://www.ipcc.ch/pdf/climate-changes-1995/ipcc-2nd-assessment/2nd-assessment-en.pdf.
7. Boucher O, Friedlingstein P, Collins B, Shine KP. The indirect global warming potential and global temperature change potential due to methane oxidation. Environ Res Lett. 2009;4:1–5. http://dx.doi.org/10.1088/1748-9326/4/4/044007
8. United Nations Framework Convention on Climate Change (UNFCCC). New York: United Nations; 1992. Available from: http://unfccc.int/resource/docs/convkp/conveng.pdf.
9. Intergovernmental Panel on Climate Change (IPCC). IPCC fourth assessment report (AR4). Working Group 1: The physical science basis. Geneva: IPCC; 2007. Available from: http://www.ipcc.ch/publications-_and_data/ar4/wg1/en/contents.html.
10. Intergovernmental Panel on Climate Change (IPCC). Climate change 2013: The physical science basis. Working Group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change. Geneva: IPCC; 2013. Available from: https://www.ipcc.ch/report/ar5/wg1/
11. De Wit M. Coal mining and carbon constraints. Final report on Task 6.1. Johannesburg: Coaltech; 2005. Available from: http://www.coaltech.co.za/.
12. Branosky E, Stevens A, Forbes S. Defining the shale gas life cycle: A framework for identifying and mitigating environmental impacts. Working paper October 2012. Washington DC: World Resources Institute; 2012.
13. Howarth RW, Santoro R, Ingraffea A. Methane and greenhouse-gas footprint of natural gas from shale formations. Clim Chang. 2011;106:679–690. http://dx.doi.org/10.1007/s10584-011-0061-5
14. Jiang M, Griffin WM, Hendrickson C, Jaramillo P, VanBriesen J, Venkatesh A. Life cycle of greenhouse gas emissions of Marcellus shale gas. Environ Res Lett. 2011;6(3):1–9. http://dx.doi.org/10.1088/1748-9326/6/3/034014
15. Cathles LM, Brown L, Taam, M, Hunter A. A commentary on 'The greenhouse-gas footprint of natural gas in shale formations' by R.W. Howarth, R. Santoro, and Anthony Ingraffea. Clim Chang. 2012;113:525–535. http://dx.doi.org/10.1007/s10584-011-0333-0
16. General Accountability Office (GAO). Federal oil and gas leases: Opportunities exist to capture vented and flared natural gas, which would increase royalty payments and reduce greenhouse gases. GAO-11–34. Washington DC: US GAO; 2010. Available from: http://www.gao.gov/new.items/d1134.pdf.
17. Forster D, Perks J. Climate impact of potential shale gas production in the EU. Report for European Commission DG CLIMA, report no AEA/R/ED57412. Didcot, Oxfordshire: AEA; 2012. Available from: http://ec.europa.eu/clima/policies/eccp/docs/120815_final_report_en.pdf
18. Broderick J, Anderson K, Wood R, Gilbert P, Sharima M, Footitt A, et al. Shale gas: An updated assessment of the environmental and climate change impacts. A report commissioned by the co-operative and undertaken by researchers at the Tyndall Centre, University of Manchester [document on the Internet]. c2011 [cited 2014 Feb 25]. Available from: http://www.tyndall.ac.uk/sites/default/files/coop_shale_gas_report_update_v3.10.pdf
19. Stephenson T, Valle JE, Riera-Palon X. Modelling the relative GHG emissions of conventional and shale gas production. Environ Sci Technol. 2011;45:10757–10764. http://dx.doi.org/10.1021/es2024115
20. Lechtenböhmer S, Dienst C, Fischedick M, Hanke T, Langrock T, Assonov SS, et al. Greenhouse gas emissions from the Russian natural gas export pipeline system. Wuppertal/Mainz: Wuppertal Institute and Max Planck Institute; 2005.
21. Eskom. Annual report 2012. Johannesburg: Eskom; 2012. Available at: http://www.eskom.co.za
22. Bradbury JA, Obeiter M, Draucker L, Wang W, Stevens A. Clearing the air: Reducing upstream greenhouse gas emissions from US natural gas systems. WRI working paper. Washington DC: World Resources Institute; 2013. Available from: http://pdf.wri.org/clearing_the_air_full_version.pdf.
23. Skone TJ, Littlefield J, Marriott J. Life cycle gas inventory of natural gas extraction delivery and electricity production. Final report 24 Oct 2011 (DOE/NETL-2011/1522). Pittsburgh, PA: US Department of Energy, National Energy Technology Laboratory; 2011.

Postgraduate research methodological flaws detected at final examination stage: Who is to blame?

AUTHOR:
Aceme Nyika[1]

AFFILIATION:
[1]Faculty of Health Sciences, University of the Witwatersrand Medical School, Johannesburg, South Africa

CORRESPONDENCE TO:
Aceme Nyika

EMAIL:
Aceme.Nyika@wits.ac.za

POSTAL ADDRESS:
Faculty of Health Sciences, University of the Witwatersrand Medical School, 7 York Road, Parktown 2193, South Africa

DATES:
Received: 11 Sep. 2013
Revised: 31 Oct. 2013
Accepted: 13 Nov. 2013

KEYWORDS:
postgraduate research; research proposal; proposal evaluation; final examination; methodological flaws

HOW TO CITE:
Nyika A. Postgraduate research methodological flaws detected at final examination stage: Who is to blame? S Afr J Sci. 2014;110(3/4), Art. #2013-0288, 4 pages. http://dx.doi.org/10.1590/sajs.2014/20130288

In the wake of globalisation, most universities are intensifying efforts to improve their overall performance in order to attract students and enhance chances of securing competitive funding from various sources. As part of these efforts, universities are striving to ensure that their postgraduate programmes meet nationally and internationally acceptable standards. Research projects conducted by students form a critical component of most postgraduate programmes and universities have put in place procedures meant to ensure that postgraduate research meets acceptable minimum standards. The procedures include setting minimum entry educational qualifications, supervision by qualified members of university academic staff, institutional evaluation of research proposals before the proposed research is embarked on, submission of progress reports by postgraduate students during the course of their programmes, and final examination of students' theses, dissertations or research reports by internal as well as external examiners. In some instances, the examiners recommend outright rejection of the student's write-up if they consider the methodology used to be inappropriate to answer research questions of the project conducted. The implications of research methodological shortcomings which are identified at the final examination stage, even if the research proposals were evaluated and approved by appropriate university structures before commencement of the research projects, are discussed. As postgraduate programmes are meant to nurture a competent and resourceful workforce and future researchers, universities need to pay attention to the issue of research methodology and internal evaluation systems in order to minimise chances of compromising the quality of their postgraduate degree programmes.

Introduction

Globally, universities aim to provide quality education which is demand driven and is relevant to dynamic needs of countries, continents and the world. In light of these aspirations, various systems of evaluating and ranking universities have emerged, with some using indicators based on the main activities of universities which are teaching, research and service provision[1-3] while others are based on cybermetric indicators.[4-7] Regardless of the system of ranking used, universities in Africa have very low scores in the global ranking system, with the top 100 universities being in developed countries. The best African universities according to the global ranking systems are in South Africa. According to The Times Higher Education World University Rankings, the best African universities are the University of Cape Town (UCT), the University of the Witwatersrand (Wits), Stellenbosch University (SU) and the University of Kwazulu-Natal (UKZN), which were at position 113, in category 226-250, in category 251-275, and in category 351-400, respectively, in the global ranking system for the year 2012.[1]

Some countries have developed their own national ranking systems which are to a large extent based on the global ranking systems. For instance, South Africa has its own ranking system which refers to the category of the best universities in the country as the 'red cluster', which in 2012 was composed of five universities: Wits, UKZN, SU, UCT and the University of Pretoria (UP).[8] With the exception of UP, these are the same African universities ranked highly in the global ranking system relative to other universities on the African continent.

It has been observed that, generally, after colonisation most universities in African countries did not develop at the same rate as universities in developed countries.[9] In some African countries, universities were initially established as part of a main parent university in the respective colonising countries, but the established African universities gradually became autonomous and independent. In addition, governments and other players such as churches and the private sector have become increasingly involved in the establishment of universities in most African countries. The slow development of the African universities could arguably be attributed to poor infrastructure[10] and inadequate resources as a result of a combination of limited injection of resources by the respective African governments and a shift in focus of foreign aid from tertiary education to primary and secondary education.[11]

Regardless of positions occupied in national or international ranking systems, universities have developed internal systems which are meant to ensure that the quality of education and training they provide meets certain minimum acceptable standards. One of the most critical roles of universities is to run postgraduate programmes, which to a large extent enhance the productivity of universities in terms of research outputs. Consequently, universities have put in place institutional procedures aimed at ensuring that they produce postgraduate students of the highest possible quality. The internal procedures vary from one university to another, but in general the first step is the screening of potential students at the admission stage so as to enroll candidates with the necessary educational background. After enrolment, the postgraduate students have to register annually and the registration is subject to satisfactory progress in the preceding year. If the postgraduate degree is by course work only, then progress is in terms of any courses that the student is required to take. For postgraduate programmes that require research projects to be done either in full or partial fulfillment of the degree requirements, progress in the research project is assessed from the proposal stage through the conduction of the research to the stage of writing up the results from the research for final examination.

A postgraduate student in a research-based programme has to submit a thesis, dissertation or research report to the postgraduate department for the final examination. The type of write-up to be submitted depends on the degree programme and the policy of the particular university. The four main possible outcomes of the examination process are (1) pass, (2) conditional pass subject to minor revision, (3) major revisions to be followed by re-examination, and (4) fail, which is an outright rejection of the write-up. Revisions which are to do with mistakes in the scientific writing or presentation of the findings could be corrected within a reasonable period of time. However, failure or rejection as a result of fundamental research methodological shortcomings would require the original research proposal to be corrected in light of the examiners' reports and the research project to be repeated by the student using the corrected methodology.

Currently there is a paucity of literature addressing pertinent issues surrounding postgraduate research and examination processes in universities. In this paper, the issues surrounding failure at the final examination stage because of methodological flaws in postgraduate research projects that were evaluated and approved by the university concerned in the first place, are discussed. Firstly, an overview is provided of university processes aimed at ensuring that research methodologies used by students are appropriate for their research questions and meet internationally acceptable standards. A discussion follows which covers various possible implications of rejection of a thesis, dissertation or research report by examiners at the final examination stage because research methods used are deemed to be inappropriate. Research is the main factor which strengthens the other complementary factors such as teaching, provision of services and ability to secure funding. Globally, postgraduate research programmes make significant contributions towards the research outputs of universities.

Postgraduate research proposal

A postgraduate student in a programme which involves a research project has to develop a research proposal under the guidance of at least one supervisor appointed by the university. The proposal has to explain in detail the research project to be conducted by the student. In general, the proposal has to have a title that captures the main aim of the intended study. An introduction has to put the study into context, explaining the background and the research question or hypothesis to be addressed by the study. A literature review should show what is already known about the research issue and the nature of any gap in knowledge which the study intends to address.

The introduction and literature review should be followed by the main objective as well as specific objectives of the study. It is the specific objectives that form the basis for the research methods which will be used. The research methodology should be explained in detail to enable readers to assess whether or not it is appropriate for the research questions to be answered. The overall research design has to be explained. The proposal should explain whether it will be a retrospective or prospective study and whether it will be qualitative or quantitative. Details such as the targeted population, sampling technique, inclusion and exclusion criteria, variables to be measured, laboratory tests to be done and statistical analyses to be done should be fully explained in the proposal if applicable. Also to be included in the research proposal are pertinent ethical issues, the work plan, budget and references.

Institutional procedures for postgraduate research

Postgraduate supervision

University students enrolled in postgraduate programmes that include research as partial or full fulfillment of the requirements of their respective postgraduate degrees must have supervisors appointed by the relevant postgraduate department in order to guide the students during the course of their programmes. Each university has its own procedures for the nomination and appointment of supervisors for postgraduate students. However, in order to foster effective supervision, universities generally require supervisors to have relevant and adequate educational qualifications and experience which are commensurate with the level of the postgraduate programme to be supervised. This requirement is intended to ensure that the supervisor has the knowledge and experience needed to be able to provide effective leadership to the student. More than one supervisor may be appointed to lead a postgraduate student if it is deemed necessary.

Institutional procedures to ensure an acceptable quality of postgraduate research start with guidance provided by the supervisor during the development of the proposal. Postgraduate students are supervised throughout their degree programme from start to finish and most universities stipulate the minimum face-to-face contact hours between student and supervisor which depend on the level of the postgraduate degree.

Postgraduate proposal

The next step in the procedure to ensure the minimum acceptable quality is the evaluation of the research proposal through an internal system involving assessors appointed by the relevant postgraduate department of the university. The evaluation of students' proposals is to ascertain the soundness of the research methodology, originality, the adequacy of the proposed work for the postgraduate degree for which the student is registered and the feasibility of the proposed study in terms of the time and funds needed. In addition to checking the scientific merits of the proposal, it is also submitted to the relevant ethics committee of the university for ethical clearance, depending on whether the project deals with humans, animals or the environment in general. The student can only start the proposed research project after obtaining clearance from the postgraduate department and from the relevant ethics committee. However, some universities exempt certain types of research projects in fields such as engineering, geology and mathematics from the requirement to undergo ethical clearance.

Progress reports

Most universities require progress reports to be submitted periodically (every semester or annually) so as to monitor the progress of the student. If progress is deemed to be unsatisfactory, appropriate remedial measures may be taken by the postgraduate department in consultation with the supervisor(s). The reports also enable the university to ensure that there is no deviation from the approved proposal. Any major methodological problems should be detected during the course of the programme as the progress reports should cover results obtained and methods used to gather the data.

Final examination

After completing the research project, the student writes a thesis, dissertation or research report to be submitted to the postgraduate department for final examination. The exact submission procedure and requirements depend on specific policies of the particular universities. The submitted postgraduate write-up is then examined by internal examiners plus external examiners appointed by the postgraduate department. The majority of universities provide institutional guidelines to be used by the postgraduate examiners, who should not have any conflict of interest. In general, the guidelines cover aspects of the research such as originality, introduction, literature review, research methodology, research results, discussion and conclusion to be assessed. Also to be assessed are the overall flow of ideas and presentation of findings in the write-up, grammar, typography and spelling. Based on the final write-up, the examination process should aim to assess the level of intellectual grasp of the concept of research and how it should be conducted, which makes research methodology a very critical aspect of postgraduate research programmes.

The examiners write reports which give their recommendations regarding the thesis, dissertation or research report which they evaluated. There are generally four possible examination outcomes that could be given by the examiners, namely (1) pass without any revisions, (2) conditional pass subject to minor revisions, (3) major revisions followed by re-examination and (4) fail without any opportunity for revision. Some universities conduct oral examinations, which are generally referred to as 'vivas',[12] after examination of the write-up. The examiners explain the basis of their decisions in their reports which are then made available to the supervisor(s) and student in order to enable the student to do any revisions which may have been deemed necessary.

Implications of research methodological flaws

Whereas other types of shortcomings could be considered to be 'peripheral', research methodology is arguably the most important aspect of research which postgraduate students, or any researcher for that matter, should understand before embarking on any research project. It is mainly the research methodology which determines whether or not the research question(s) can be answered. Hence most universities have systems in place which evaluate research proposals before postgraduate students can begin their research projects.

It follows therefore that having a postgraduate research thesis, dissertation or research report being condemned by examiners on the basis of flawed research methodology raises several issues and questions which need to be addressed if the quality of postgraduate research in universities is to be enhanced. The scenario of research methodological flaws being detected at the final examination stage has four main implications, all of which reflect badly on the university concerned.

First implication: The proposal was flawed

The first main implication is that the research proposal which was evaluated by the university postgraduate system in the first place was flawed but the internal evaluation system did not detect the research methodological shortcomings. The internal system includes the supervisor(s) who guided the student during proposal development and the postgraduate structures, such as the departmental and faculty higher degrees committees, which were responsible for evaluating and approving the postgraduate research proposal. The possibility that flawed proposals can be approved and research projects completed before shortcomings are detected at the last stage of the postgraduate programme, brings into question the effectiveness of the whole evaluation system in ensuring high-quality postgraduate research. Such a scenario implies that the steps in the internal system are carried out as mere formalities without rigorous evaluation of research proposals.

Second implication: The proposal was inadequate

The second implication is that the format of the proposal does not enable adequate evaluation of the research methodology. This situation would exist if the proposal was so abridged that insufficient detail about the research methodology was included. Hence such a proposal could sail through the internal evaluation system without any methodological flaws being detected, and the flaws would become evident when the detailed thesis, dissertation or research report is submitted for examination. Some universities require that the postgraduate student presents and 'defends' the proposal in the postgraduate department so that proposal assessors can ask for details and explanations which may not be clear in the proposal.

Third implication: The proposal was deviated from

The third main implication is that the postgraduate student deviated from the assessed and approved research proposal during the course of the programme and ended up with data collected through inappropriate research methods. Consequently, the collected data, the analyses, interpretation of findings and conclusions contained in the thesis, dissertation or research report would not be appropriate to answer the research question(s) as explained in the approved proposal. However, such a deviation would imply that the checks and balances put in place by the university to ensure that postgraduate students are guided in the correct path of research are not effective. It would mean that the supervision process either could not detect the deviation as it was occurring, or the deviation was detected and deemed to be appropriate.

If the deviation was not detected, it implies that several qualified and experienced members of the academic staff of the university concerned would have failed to pick up the major methodological shortcomings which were eventually detected by examiners at the final examination stage of the programme. If the deviation was actually allowed by the supervisors and it is eventually criticised by examiners as being inappropriate for the research, then the question arises as to who is correct – the supervisors or the examiners. It would be unfair, unethical and unprofessional to blame and penalise the postgraduate student who is an unqualified and inexperienced researcher enrolled in the postgraduate programme to gain the requisite qualification and experience.

Fourth implication: The examiners are wrong

The fourth main implication is that the examiners are wrong. In other words, there are no research methodological shortcomings in the final thesis, dissertation or research report but the examiners incorrectly think that there are shortcomings. If all the examiners independently and separately, but incorrectly, criticise the research methodology, then one implication is that the criteria of selecting and appointing examiners is not effective. If there are discordant reports submitted by examiners, then some of them may be wrong while others may be right, but it becomes a challenge to determine who is wrong and who is right. Another possible implication of this scenario is that the guidelines developed by the university to guide the examiners are not effectively providing the intended guidance and marking framework, which could lead to some examiners making flawed judgements.

Recommendations to consider

A major recommendation to consider is for universities to review their postgraduate evaluation systems to ensure that they effectively serve the purpose for which they were put in place. If there are cases of thesis, dissertations or research reports criticised at the final examination stage as a consequence of methodological problems, the university should investigate the cases starting with re-evaluation of the original proposals to determine if flaws slipped through without detection. The supervision process, including progress reports submitted and minutes of meetings between supervisor and students, should be assessed, preferably by assessors independent of the university. Finally, the condemned thesis, dissertation or research report should be re-examined by independent examiners with relevant qualifications and experience. The re-examination reports should then be compared with the 'controversial' examiners' reports. Such a thorough and comprehensive investigation would enable the university to assess how issues of research methodology could crop up at the final examination stage after the research proposals and research processes had been subjected to internal evaluation systems. Any loopholes which may be identified in the system should then be addressed and the system should be assessed regularly thereafter.

Another recommendation to consider is to improve (if already in existence) or put in place support mechanisms aimed at strengthening postgraduate supervision in the university. Elementary and advanced training workshops could be developed to cater for emerging and experienced supervisors, respectively. As for the examination process, it would be important to assess the guidelines for examiners to be sure that they are up to date and clearly conform with the policy of the university regarding postgraduate programmes that involve research. It would also be helpful to increase the pool of internal and external examiners from which to choose appropriate examiners for specific types of research. The increase should be both in terms of the number of examiners and the diversity of areas of specialisation. One possible method to increase the pool of examiners is to offer reasonable honoraria in order to attract and retain desired examiners. Universities could also consider developing and running appropriate training workshops for examiners so as to ensure uniform institutional marking schemes and standards pertaining to postgraduate research. In addition, the issue of weighting of the internal versus external examiners' comments and marks should be dealt with and clarified in the university's policy or guideline documents.

Conclusion

Postgraduate research programmes contribute significantly towards the overall productivity of universities and the quality of their products. The advancement of technology in various fields depends to a large extent on the continuous availability of resourceful and innovative researchers for the future. Universities across the world strive to collectively produce such researchers through postgraduate research programmes. It is therefore critical that universities take measures to ensure that their postgraduate research programmes effectively impart the required profound intellectual comprehension of research methodologies.

Research methodology queries arising at the final examination stage of postgraduate programmes raises questions about the quality and appropriateness of the evaluation systems of the universities concerned. It arguably implies ineffectiveness on the part of the student, the postgraduate supervisors, postgraduate proposal assessors and or examiners. Regardless of where the blame actually lies, emergence of research methodological shortcomings at the final examination stage of a postgraduate thesis, dissertation or research report gives a bad impression of the university concerned. Overall, universities demonstrating high productivity and internationally acceptable standards attract the best students, the best workforce and sustainable funding opportunities.

References

1. Times Higher Education World University Rankings 2012–2013 [homepage on the Internet]. No date [cited 2013 Aug 19]. Available from: http://www.timeshighereducation.co.uk/world-university-rankings/2012-13/world-ranking
2. Academic ranking of world universities [homepage on the Internet]. No date [cited 2013 Aug 20]. Available from: http://www.arwu.org/
3. QS university ranking methodology [homepage on the Internet]. No date [cited 2013 Aug 20]. Available from: http://www.iu.qs.com/university-rankings/world-university-rankings/
4. Aguillo IF, Granadino B, Ortega JL, Prieto JA. Scientific research activity and communication measured with cybermetrics indicators. J Am Soc Inf Sci Technol. 2006;57(10):296–302. http://dx.doi.org/10.1002/asi.20433
5. Ortega JL, Aguillo IF. Mapping world-class universities on the web. Inform Process Manag. 2009;45:272–279. http://dx.doi.org/10.1016/j.ipm.2008.10.001
6. Aguillo I. Measuring the institution's footprint in the web. Library Hi Tech. 2009;27(4):540–556. http://dx.doi.org/10.1108/073788309
7. Ranking Web of universities [homepage on the Internet]. No date [cited 2013 Aug 20]. Available from: http://www.webometrics.info/en/Methodology
8. Rankings of universities in South Africa [homepage on the Internet]. No date [cited 2013 Aug 22]. Available from: http://en.wikipedia.org/wiki/Rankings_of_universities_in_South_Africa
9. Gukas I. Global paradigm shift in medical education: Issues of concern for Africa. Med Teach. 2007;29:887–892. http://dx.doi.org/10.1080/01421590701814286
10. Burdick W. Challenges and issues in health professions education in Africa. Med Teach. 2007;29:882–886. http://dx.doi.org/10.1080/01421590701821695
11. Bloom D, Canning D, Chan K. Higher education and economic development in Africa [document on the Internet]. c2006 [cited 2013 Sep 10]. Available from: http://www.sciencedev.net/Docs/Higher%20Education%20and%20economic%20developmnet.pdf
12. Carter B, Whittaker K. Examining the British PhD viva: Opening new doors or scarring for life? Contemp Nurse. 2009;32:169–178. http://dx.doi.org/10.5172/conu.32.1-2.169

Palaeomagnetic and synchrotron analysis of >1.95 Ma fossil-bearing palaeokarst at Haasgat, South Africa

AUTHORS:
Andy I.R. Herries[1]
Peter Kappen[2,3]
Anthony D.T. Kegley[4]
David Patterson[3]
Daryl L. Howard[3]
Martin D. de Jonge[3]
Stephany Potze[5]
Justin W. Adams[6]

AFFILIATIONS:
[1]Australian Archaeomagnetism Laboratory, Archaeology, Environment and Community Planning, La Trobe University, Melbourne, Australia
[2]Physics, La Trobe University, Melbourne, Australia
[3]Australian Synchrotron, X-ray Fluorescence Microprobe Beamline, Melbourne, Australia
[4]Biomedical Sciences, Grand Valley State University, Allendale, MI, USA
[5]Plio-Pleistocene Palaeontology Section, Ditsong National Museum of Natural History, Pretoria, South Africa
[6]Anatomy and Developmental Biology, Monash University, Melbourne, Australia

CORRESPONDENCE TO:
Andy Herries

EMAIL:
a.herries@latrobe.edu.au

POSTAL ADDRESS:
Archaeology, Environment and Community Planning, La Trobe University, Martin Building, Melbourne Campus, Bundoora, Melbourne 3086, Victoria, Australia

DATES:
Received: 11 Apr. 2013
Revised: 17 Aug. 2013
Accepted: 09 Sep. 2013

KEYWORDS:
biochronology; X-ray flourescence; synchrotron radiation; iron; primate fossils; magnetostratigraphy

Palaeomagnetic analysis indicates that Haasgat, a fossil-bearing palaeocave in the Gauteng Province of South Africa, is dominated by reversed magnetic polarity in its oldest, deepest layers and normal polarity in the younger layers. The presence of in-situ *Equus* specimens suggests an age of less than ~2.3 Ma, while morphological analysis of faunal specimens from the ex-situ assemblage suggests an age greater than 1.8 Ma. Given this faunal age constraint, the older reversed polarity sections most likely date to the beginning of the Matuyama Chron (2.58–1.95 Ma), while the younger normal polarity deposits likely date to the very beginning of the Olduvai Sub-Chron (1.95–1.78 Ma). The occurrence of a magnetic reversal from reversed to normal polarity recorded in the sequence indicates the deposits of the Bridge Section date to ~1.95 Ma. All the in-situ fossil deposits that have been noted are older than the 1.95 Ma reversal, but younger than 2.3 Ma. Haasgat therefore dates to an interesting time period in South African human evolution that saw the last occurrence of two australopith species at ~2.05–2.02 Ma (Sts5 *Australopithecus africanus* from Sterkfontein Member 4) to ~1.98 Ma (*Australopithecus sediba* from Malapa) and the first occurrence of early *Homo* (Sk847), *Paranthropus* and the Oldowan within Swartkrans Member 1 between ~2.0 Ma and ~1.8 Ma.

Introduction

The exposed Malmani dolomite to the west of Johannesburg and Pretoria in South Africa is the host rock for thousands of cave systems that have been forming throughout most of the Quaternary (last 2.6 Ma) and perhaps as early as the beginning of the Pliocene (5.3 Ma). These cave systems are well known for their wealth of archaeological and fossil-bearing infills, including abundant hominin remains.[1-3] While a large body of research exists on the hominin-bearing palaeocave deposits such as Sterkfontein[4-6] and Swartkrans[7], recent results from less intensively worked or new fossil localities, such as Malapa[8-11], the Bolt's Farm complex[12,13], Gondolin[1,14-16] and Hoogland[17], have highlighted the importance of wider regional sampling for understanding geographical and temporal variability in Plio-Pleistocene karst development, faunal evolution, taphonomy and palaeoecology that underlie interpretations of the South African hominin and primate fossil record.

Unfortunately, an historical inability to reliably date these fossil localities has limited attempts to integrate the data from South African palaeocaves, as well as to compare the data to the record from eastern African Plio-Pleistocene sites. Some of the earliest attempts at dating the Swartkrans and Makapansgat Limeworks hominin-bearing cave deposits using palaeomagnetism[18,19] produced results, but inappropriate sampling (e.g. drill cores, breccia) as well as mistakes in understanding the stratigraphy of these complicated sites, provided little clarity. Recent improvements in palaeomagnetic correlation methods for South African karstic systems[6,9,10,15,17,20,21] and integration of uranium-lead dating[5,9,10] with electron spin resonance[1,6] have begun to unravel the complex history of these sites; for some deposits, the application of multiple methods has provided the first reliable age estimates[1-3].

The Haasgat cave system is a little explored fossil-bearing palaeocave that lies in the Monte Christo Formation of the Malmani dolomites that forms part of the Schurveberg Mountain Range west of Pretoria, ~20 km northeast of the well-described early Pleistocene karst systems of Sterkfontein, Swartkrans and Kromdraai (Figure 1). Haasgat is located on the Leeuwenkloof 480 JQ farm, positioned between the Gondolin hominin-bearing palaeocave (~4 km to the northeast)[1,14-16] and the Malapa and Gladysvale australopith-bearing palaeocaves (~5–7 km to the southwest).[8-11,21] Limited geological investigation and ex-situ palaeontological sampling in the late 1980s produced a relatively large and diverse sample of extinct baboon, colobus monkey and ungulate species.[22-26] In 2010, geological and palaeontological research was renewed at Haasgat with the aim of more reliably determining the age of the deposit, conducting new excavations of the first in-situ deposits, and undertaking a thorough reanalysis of the previously recovered Haasgat HGD fossil sample.[22] Here we present the first magnetobiostratigraphic analysis of the Haasgat fossil locality.

HOW TO CITE:
Herries AIR, Kappen P, Kegley ADT, Patterson D, Howard DL, De Jonge MD, et al. Palaeomagnetic and synchrotron analysis of >1.95 Ma fossil-bearing palaeokarst at Haasgat, South Africa. S Afr J Sci. 2014;110(3/4), Art. #2013-0102, 12 pages. http://dx.doi.org/10.1590/sajs.2014/20130102

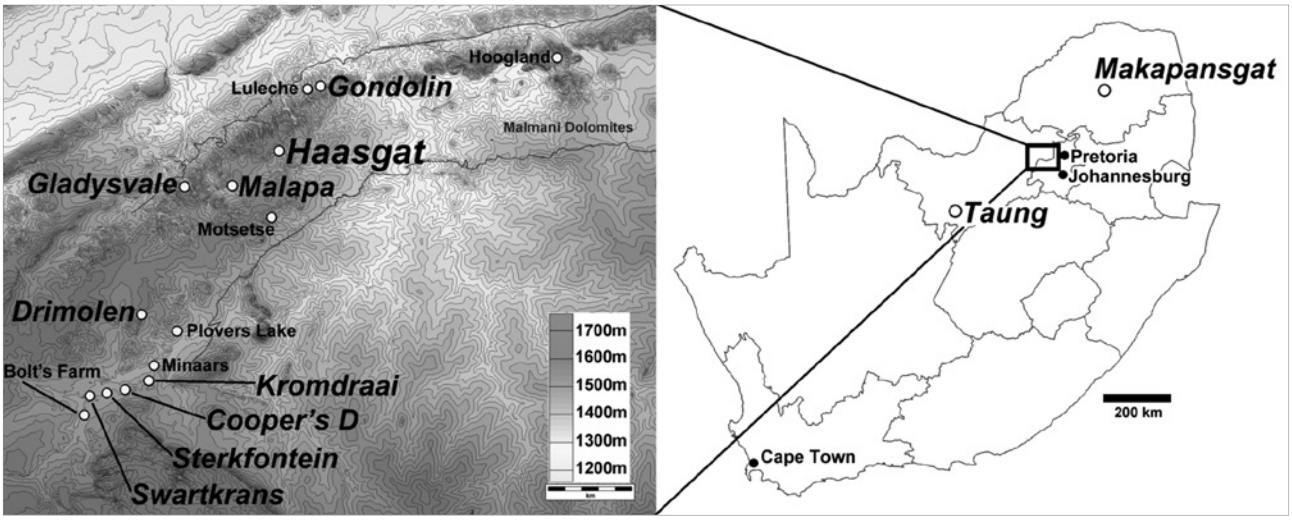


Figure 1: The location of Haasgat within the Cradle of Humankind World Heritage site and its relation to other fossil sites mentioned in the text. Sites with larger font are hominin-bearing sites (except Haasgat) and those with smaller font are vertebrate fossil sites.

History of work at the site, geology and biochronology

Palaeontological deposits in the Haasgat karstic system were first noted in 1987 and led to initial geological description[23] and faunal analysis[24-26] of fossils processed from the extensive ex-situ miner's rubble at the site. Lime mining during the early 20th century removed a large portion of flowstone, stalagmites and stalactites that formed over a basal collapse and left extensive fossiliferous bands of calcified sediments on the walls and ceiling of the mine. In removing these, the miners broke through a wall of speleothem at the end of the fossil tunnel and into the bottom of a more recent shaft that is choked a short distance into another small cave at the surface. The site is unusual compared with many of the fossil-bearing palaeocaves in the region in that the current deposit retains the original structure (roof, walls and floor) of the palaeocave, consisting of a long tunnel that was completely filled to the roof with breccia, conglomerate and fine-grained laminated sediments and speleothem. During the early palaeontological work an unknown quantity of ex-situ breccia from the associated dumpsite was collected and processed, yielding a diverse faunal sample (HGD assemblage) that includes the largest and most demographically diverse accumulations of two primate taxa (*Papio angusticeps* and *Cercopithecoides haasgati*) from a single locality in South Africa.[23-28]

Early publications on Haasgat[23,24] suggested that the elevation of the system relative to the modern valley floor, the erosional deroofing prior to mining and the occurrence of the extinct primate *Parapapio broomi* might indicate a terminal Pliocene age for primary fossil deposition (then about 1.8 Ma). Later the entire papionin sample was reclassified to *Papio angusticeps*[25] and, given the shared occurrence of the species, the age estimate for the site was considered to be contemporaneous with Kromdraai A/B and Cooper's. Kromdraai B is now dated to between 1.8 Ma and ~1.6 Ma[1,2,27], while Cooper's D is dated to between 1.6 Ma and <1.4 Ma[29]. However, no dates exist for the original Cooper's Cave sites on which this correlation was based or for Kromdraai A. Subsequently, Plug and Keyser[26] described an essentially modern ungulate sample and suggested that the deposits maximally formed between 1.5 Ma and 0.5 Ma.

Our work with the HGD assemblage since 2010 has revealed that prior publications considered only a subset (1475 of 2413) of the originally processed ex-situ specimens. Incorporation of the previously undocumented specimens into the HGD assemblage and morphological analysis of the HGD fossils have significantly altered both the faunal list and the biostratigraphic implications of the ex-situ sample.[22] Adams[22] concluded from the data that the ex situ faunal sample likely dated to sometime between ~2.3 Ma and ~1.9 Ma. Given new dates for the first occurrence of the genus *Equus* in Africa in the Omo Shungura lower Member G deposits,[30] the recovery of two species of *Equus* (*Equus capensis* and *Equus* cf. *quagga* [*sensu*[30]]) in the HGD ex-situ sample[22] indicates at least some of the faunal assemblage is younger than 2.31 Ma. (Tuff G is now dated to 2.27±0.04 Ma and the *Equus* fossils are in deposits G1 that are younger than this age, but older than 2.19±0.04 Ma – the age for Tuff G3.[31]) Although neither of the suid craniodental remains were specifically diagnostic, they express crown morphology more advanced than *Metridiochoerus* from Makapansgat Limeworks Member 3 (2.85–2.58 Ma[3,32,33]) but are more primitive relative to *Metridiochoerus* and *Phacochoerus* specimens recovered from later Pleistocene deposits like Cornelia-Uitzoek (1.07–0.99 Ma)[22,34] and overall resemble *Metridiochoerus andrewsi* specimens from Gondolin GD 2 (1.95 Ma to ~1.78 Ma[1,14,15,35]) and *M. andrewsi* and *Phacochoerus* specimens from Swartkrans Member 1 (dating to sometime between 2.25 Ma and 1.80 Ma[7], but most likely <1.96 Ma[1]). The most common bovid species in the HGD assemblage is klipspringer (*Oreotragus* sp.; 38.0% of the bovid assemblage and 15.6% of the total assemblage), which exhibits greater affinity to the sample of the genus from the Makapansgat Limeworks Member 3 deposits (2.85–2.58 Ma) than to that from Gondolin GD 2 (1.95 Ma to ~1.78 Ma[14,15,35]), despite the geographic proximity of Gondolin to Haasgat.[22] As noted above *P. angusticeps* (24.0% of the total Haasgat HGD assemblage) appears to represent an early Pleistocene papionin, although the first and last appearance dates of the species is unknown.[27] And, finally, within the colobine sample (11.0% of the total assemblage), some of the craniodental specimens attributed to the new species *Cercopithecoides haasgati*[28], but recently revised as representing *Cercopithecoides williamsi*[22,36], show some morphological similarities to specimens from Sterkfontein Member 4 (e.g. SWP 495 partial skull and SWP 1735 mandible; 2.6–2.0 Ma[3,5,6]). Collectively, these taxa represent nearly 50% of the identifiable HGD specimens and suggest an age between 2.3 Ma and 1.8 Ma for some of the ex-situ faunas; this time range is useful for magnetostratigraphic analysis because of the occurrence of a number of geomagnetic field reversals and events that have been documented at other South African palaeocave sites (Olduvai event, 1.95–1.78 Ma; pre-Olduvai event, ~1.98 Ma; Huckleberry Ridge event, ~2.04 Ma; and Réunion event, ~2.16 Ma[3,6,9,10,14]).

Detailed site description and sampling

The geology of the site has been described previously[23] and its features, plan and long section are shown in Figures 2–4. The original Haasgat cave appears to have been a long horizontal passage, likely once connected to palaeokarst remnants that occur at the same level along the margins on both sides of the steep valley in which the cave is located. This likely connection indicates that the entire north–south valley itself was once occupied by a now collapsed Pliocene (or earlier) cave system, of which Haasgat was simply one extension. A similar situation has been envisaged in the mountainous karst at Makapansgat.[37-39] The sediment infilling the main Haasgat passage today is not related to this earlier phase of the cave's life history but to a later phase of breakdown and decay during which the tunnel was truncated by the collapse of the cave and formation of the valley. This episode of erosion and downcutting would have opened the cave to the surface and begun the deposition of sedimentary deposits and fossils. However, this was not the last phase of karstification of the deposit, as shown by the more recent shaft at the end of the fossil tunnel, as well as the remnants of later, much smaller cave passages that formed through the earlier palaeocave fill. A very limited number of fossils has been noted within the fill of these later tunnels and such features could explain the occurrence of what seem like much younger fossils within the ex-situ assemblage,[22] such as *Damaliscus dorcas* and *Connochaetes gnou* (current first appearance date at Cornelia-Uitzoek: 1.07–0.99 Ma[34]). *D. dorcas* is known from Elandsfontein, although these deposits are themselves undated by radiometric methods and have an age estimate of ~1.1–0.6 Ma based on faunal correlation with South African sites and preliminary palaeomagnetism.[40,41] Moreover, this species has previously been noted at Swartkrans Member 2 (1.7–1.1 Ma[35,42]) and the genus is recorded at several sites post 2 Ma (e.g. Swartkrans Member 1[35]), but the specimens have not been designated to species level. These species are represented by only a handful of fossils (NISP 8[21]) out of the 1446[22] identifiable specimens, at least 50% of which suggest an earlier time period (see above). The site has had only extremely limited secondary karstification (one later and limited phase that is infilled with uncalcified dark brown sediment that is distinct from the red to reddish brown siltstones and mudstones of the older phase), unlike complex multi-generational sites such as Sterkfontein and Swartkrans where serious mixing is likely to have occurred between older and younger Members.[1] As such, we consider the bulk of the ex-situ fauna to come from deposits equivalent to the single phase of karstification being dated in this analysis, especially given that the main bone beds have been identified within the interior, older part of the cave.

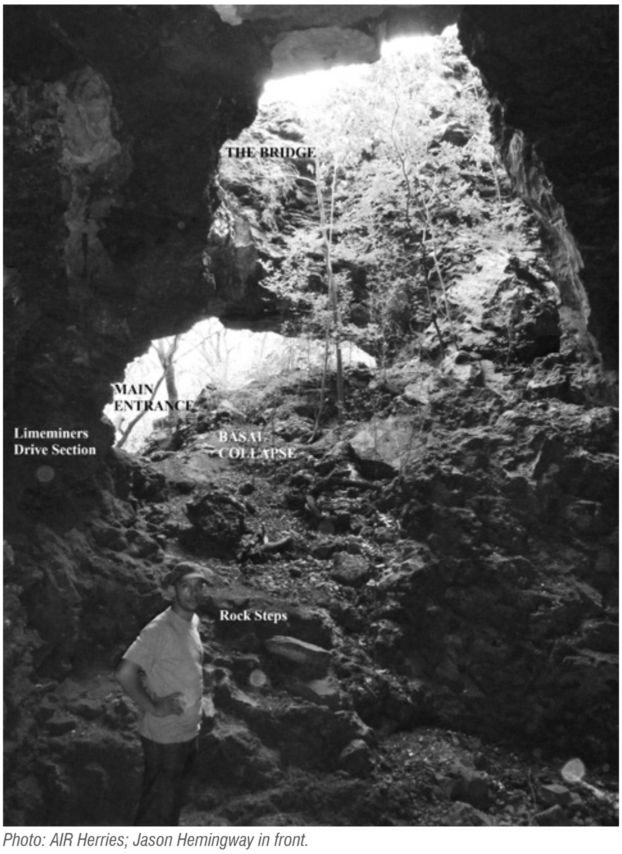


Photo: AIR Herries; Jason Hemingway in front.

Figure 2: Major features of the Eastern Deposits of the Haasgat mine and its palaeocave deposits looking east from the entrance to Tetley's Hall (perspective shown in plan of Figure 4).

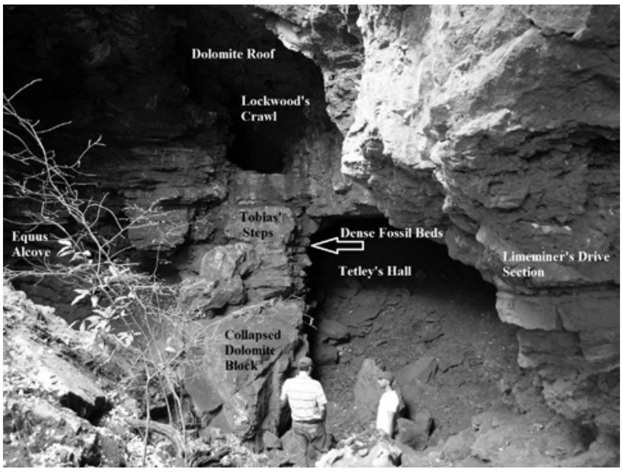
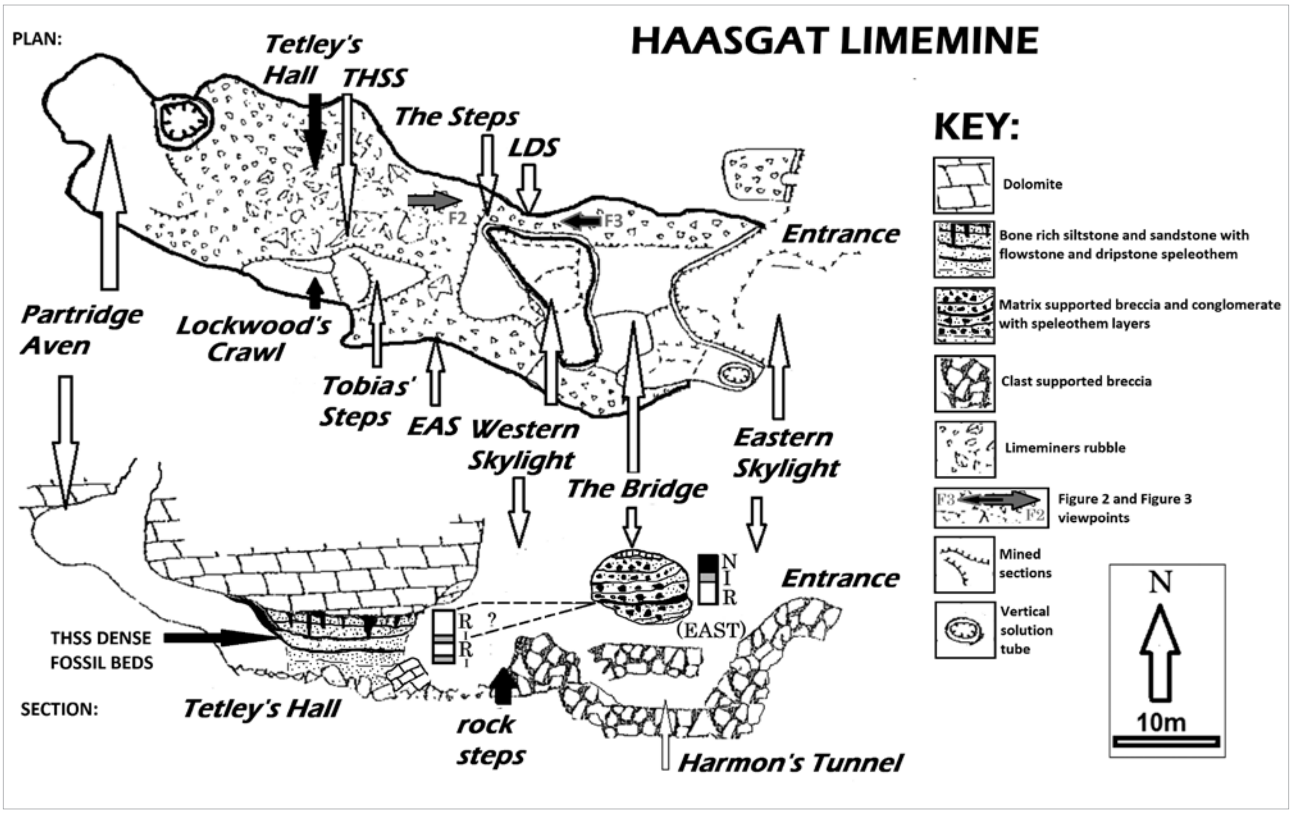


Photo: AIR Herries; Jason Hemingway and Colin Menter can be seen in the photo.

Figure 3: Major features of the Western Deposits of the Haasgat mine and its palaeocave deposits looking west from beneath the Bridge (perspective shown in plan of Figure 4).

Today, Haasgat consists of a large mined tunnel running almost east–west into the hillside with a large entrance at its easterly end. The current cavity is almost entirely the result of lime mining activity for speleothem which has removed a number of generations of cave infill in varying stages of calcification. Unlike many palaeocave fills of this type, in which the original cave has been completely infilled with sediment and partly eroded away, the roof and walls have been preserved in the rear section of the tunnel, while the floor remains buried under both the original basal cave rubble and subsequent mining rubble. The mined tunnel ends at a slope up into the bottom of a high chocked shaft (Partridge Aven) that has been infiltrated by tree roots from vegetation growing in the small cave above the fossil site (Keyser Pot). Before the lime mining there would have been no access horizontally into this part of the system as the current route would have been filled with heavily calcified sediment and speleothem. It appears that Partridge Aven was at some time at least partially filled by sediments as shown by the sediment remnants adhering to its upper reaches of the Aven as viewed from below. How Keyser Pot and Partridge Aven relate to the main palaeokarstic fill is uncertain, but the impression is that it is a much younger fill from a later phase of karstification related to vertical, rather than horizontal, development. The reuse of palaeokarstic conduits by water and the development of new cave passages within old palaeokarst is a common feature of caves in this part of South Africa and can also be seen in other South African dolomite areas such as at the Sudwala Caves (Herries personal observation). This situation is in part related to the fact that calcified palaeokarst contains purer calcite and is thus more easy to dissolve than the dolomite.

A number of morphological features exist in the current cavity (mine) that aid understanding the magnetobiostratigraphic interpretation and they will be briefly described based on a person entering from the main eastern entrance to the mine (Figures 2–4). At the entrance to the mine is a large platform of rubble from lime mining operations into the Haasgat palaeocave deposits. This mining talus cone, which extends some 30 m down the hillside to the valley floor, appears to have served as the source for at least some of the 1988 HGD ex-situ fossil sample. Palaeocave fill outcrops directly onto the hillside, and prior to mining the outline of the fossil tunnel would have been visible after vegetation was cleared. It is evident that the lime miners began by creating an east–west drive through the palaeocave fill and along the very northern wall of the tunnel. This drive also cut into the heavily calcified basal dolomite and chert collapse rubble (basal breccia) that formed as the tunnel developed as a result of the roof collapse. Despite subsequent mining, the outline of this drive is still preserved. The drive seems to have stopped at the approximate position of what is now a buttress of palaeocave remnants adhering to the north wall of the cave and has been termed the 'Limeminer's Drive Section' (LDS). This buttress consists of a series of



THSS, Tetley's Hall Southern Section; LDS, Limeminer's Drive Section; EAS, Equus Alcove Section

Polarity: N, normal; I, Intermediate; R, reversed

Figure 4: A plan view and a cross-sectional view of the Haasgat Limemine (redrawn using data of Keyser and Martini[22]) showing the major features and stratigraphy described in the text and composite palaeomagnetic polarity of the western and eastern sections. The location of the main in-situ bone bed is shown, as are the perspectives of Figures 2 and 3.

interstratified speleothem and sandstone at its base, grading up into finer grained siltstone that connects with sediments outcropping in the roof at the rear of the cavity. As a result of its isolation, the sediments in this buttress were not sampled in this initial study.

The Limeminer's Drive also passes under a large bridge of palaeo-sediment fill (the Bridge Section East/West; BSE/BSW) that is connected to both the northern and southern walls of the tunnel (Figures 2 and 3). A small remnant of the original tunnel roof is preserved at the very top of this Bridge and sampling from this roof to the base of the sediments was undertaken on both its east and west face using single rope technique.[40] This sequence consists mostly of conglomerate and breccia deposits that show that this part of the palaeosediments was closer to the entrance than those in the western sector of the tunnel. It also indicates significant waterflow through this part of the cave which would have been responsible for winnowing fine-grained material into the deeper western sections of the tunnel, which is dominated by fine-grained and well-laminated siltstone, sandstones and mudstones. Because of the more brecciated nature of the sediments in the Bridge Sections, sampling for suitable palaeomagnetic samples was more difficult, as breccia rarely records good palaeomagnetic signals. Breccia's poor signal is because it forms as a result of rapid collapse, and so the magnetic grains have no mechanism for accurate orientation and fossilisation of the field direction at the time of deposition. Moreover, the large number of inclusions in breccia causes a randomisation of the magnetic signal because each clast has its own independent magnetic polarity and direction. Fortunately, some finer-grained siltstone layers do occur throughout the sequence, as do flowstone layers, and these layers were the main target for palaeomagnetic analysis.

At its base the Bridge sequence rests on a large fallen dolomite block that sits on the top, but partly within the basal breccia that was excavated around by the limeminers. Beneath this fallen block on the eastern side of the bridge a hole leads down into a lower passage that, at least in part, has been formed through the original basal collapse breccia of the palaeocave (Harmon's Tunnel) by later karstic processes than those that formed and infilled the palaeocave. The deposits exposed in this tunnel are apparently sterile breccias and no sampling was undertaken in this study. This tunnel leads through and back up to '*Equus* Alcove' and into the tunnel's eastern section (see below). On either side of the Bridge there are two large skylights (Eastern and Western Skylights) to the surface that are the product of later opencast mining from the surface. The Western Skylight ends at the beginning of 'Tetley's Hall', the largest entirely subterranean portion of the current cavity and the point at which the original roof of the tunnel is entirely preserved. Beneath the Eastern Skylight the floor consists of two large collapsed dolomite boulders and it is at roughly this position that the Limeminer's Drive ends at a series of artificially formed rock steps that lead down to the rubble-filled modern floor of the mine.

Climbing up a slight slope from the modern floor to the left (south) from these steps brings you to a long alcove along the southern wall of the tunnel which preserves a varying thickness of fine-grained, well-laminated palaeocave fill with very thin (a few centimetres at most) interstratified flowstones. This area has been termed the '*Equus* Alcove Section' (EAS) because of an exposed in-situ *Equus* metacarpal calcified within these deposits. At the eastern end of this alcove are a series of massive mined steps of fine-grained calcified sediment (Tobias' Steps) that can be climbed into a higher tunnel near the cave roof (Lockwood's Crawl). The walls of Lockwood's Crawl consist of very finely laminated mudstone with the occasional thin flowstone layers (a few centimetres thick) and, near the exposed dolomite roof, quite dense layers of fossils. The original dolomite roof of the cave is exposed here and shows clear signs of anastomosis (the development of a network of branching, intersecting and rejoining tubes). Cave anastomoses are generally formed by dissolution of the dolomite by slow, poorly directed, phreatic flow along a bedding plane (here forming a westerly sloping roof) in shallow dipping rock. Such features represent an important process in the early stages of cave development and are a primary driver of roof collapse and the slow vertical development of cave passages that are primarily developing in the horizontal plane as a result of major bedding control and insoluble chert layering.

The sediments making up Tobias' Steps were sampled for palaeomagnetic analysis and during sampling a number of primate fossils were recovered. In an alcove at the base of these steps a complete primate mandible was also present, indicating that larger, more complete fossils occur in the fine-grained silts that could not have been washed there by winnowing processes from the entrance breccia. Tobias' Steps form the beginning of the southern wall of Tetley's Hall and the eastern end of Tetley's Hall Southern Section (THSS) and it is here that the oldest and thickest fossil-bearing sediments in the Haasgat depositional sequence are exposed. This section was also sampled for palaeomagnetic analysis; it consists of a series of fine-grained siltstone and sandstones regularly interstratified with thin flowstones and sporadically interstratified with thicker flowstones up to 0.12 m deep. The source for these flowstones is a series of smaller stalagmites and one large stalagmite boss which formed at the rear of the tunnel and is in part what separated Partridge Aven from the palaeocave sediments. The deposits at the eastern and western ends of THSS and in Tobias' Steps are well calcified, while deposits in the middle of THSS are partly decalcified, which is likely a result of the formation of a later cavity through the palaeocave fill in this area that has been mined away. The northern wall of Tetley's Hall consists almost entirely of the dolomite and chert wall of the ancient tunnel covered by a wad in places (e.g. manganese dioxide residue from the erosion of the dolomite prior to the opening up of the cave to the surface). Following the southern wall of Tetley's Hall leads to a point where the far end of Lockwood's Crawl punches through the palaeocave fill about 2.5 m above the current floor and enters what is almost the furthest point of Tetley's Hall. Beyond this point the original cave was filled with speleothem that was mined through and now forms a slope up into the base of Partridge Aven. The current floor of Tetley's Hall is made up of limeminer rubble, and palaeocave sediments can be seen disappearing beneath this rubble on the southern wall of Tetley's Hall (THSS).

The roof of Tetley's Hall is composed of sediment that also outcrops within Lockwood's Crawl (Lockwood's Crawl Section) and which can also be seen on the edge of the Western Skylight. Within the Skylight a series of flowstone layers and speleothems can be seen to have been deposited on the basal deposits that are equivalent to those in THSS and have been infilled around and capped by later sediment that outcrops both in Lockwood's Crawl and in the top of the EAS. As such there is contiguous stratigraphy between the basal deposits in Tetley's Hall (THSS) through the sediments outcropping in Lockwood's Crawl (LCS) and EAS and up into the youngest deposits exposed in the top of the Bridge Sections (BSE/BSW). However, in some cases it is difficult to estimate the exact layers that correlate because of remnants of unmined speleothem adhering to the walls and obscuring parts of the sections. Despite this, as you progress deeper into the cave from the entrance the deposits exposed in the walls generally get older, with the base of the Bridge Sections contemporary with the upper part of the LCS, EAS and THSS. These are more informally referred to as the Eastern (BSE/BSW, LDS) and Western Deposits (LCS, EAS, THSS). The relationship between the deposits in the LDS and the Western Deposits is more difficult to determine but generally the LDS deposits are older than the deposits in the Bridge Section and as such may be partly contemporaneous with the base of the Western Deposits. Work on the geology is ongoing using detailed stratigraphic mapping and micromorphological analysis that will help tie the various sections and layers together to create an accurate composite stratigraphy.

Palaeomagnetic methodology

Palaeomagnetic sampling was exclusively undertaken by block sampling with a hammer and chisel using the single rope technique for safety.[43] Because the cave environment is dark, the samples were oriented before removal with a region 4 Suunto cave surveying compass and clinometer. Work was undertaken at the Australian Archaeomagnetism Laboratory (TAAL) at La Trobe University (TAAL: www.archaeomagnetism.com) and the University of Liverpool Geomagnetism Laboratory (ULGL) with the magnetic cleaning of samples undertaken on the same equipment at both laboratories as part of an inter-laboratory comparison. The declination of the final results were then corrected to account for local secular variation using the 11th-generation International Geomagnetic Reference Field model through the British Geological Survey[44] (-14.859° declination and -55.828° inclination). Sub-samples from each layer or block were subjected to a range of magnetic demagnetisation techniques in order to isolate the characteristic remanent magnetisation (ChRM) from more recent viscous overprints. A series of sub-samples from each block or layer were subjected to a 11–17 point alternating field demagnetisation using a Molspin© (UK) or ULGL in-house-built alternating field demagnetiser, thermal demagnetisation using a Magnetic Measurements© (UK) MMTD80a and hybrid demagnetisation consisting of a series of initial alternating field demagnetisation cleaning steps between 8 and 12 millitesla (mT) followed by thermal demagnetisation.[6,9,14,20] Samples were measured on an AGICO© (Czech Republic) JR6 magnetometer with a noise of 2.4 μA/m. Sub-sample directions were defined using principal component analysis and were accepted with a median angle of deviation of <15°. Final directions for each block or layer were defined using Fisher statistics using the FISH2 program (written in house at ULGL) and normal and reversed polarity was defined based on the palaeopole direction. Samples with palaeolatitudes of <+45º or <-45° were defined as intermediate and those between +45–60º and -45–60° were defined as intermediate normal or reversed, respectively. Those with palaeopoles greater than +60° and -60º were defined as normal or reversed polarity, respectively. Mineral magnetic measurements were undertaken with a Bartington© (UK) MS3 magnetic susceptibility system at TAAL and a Magnetic Measurements© variable field translation balance at ULGL.

X-ray fluorescence microscopy

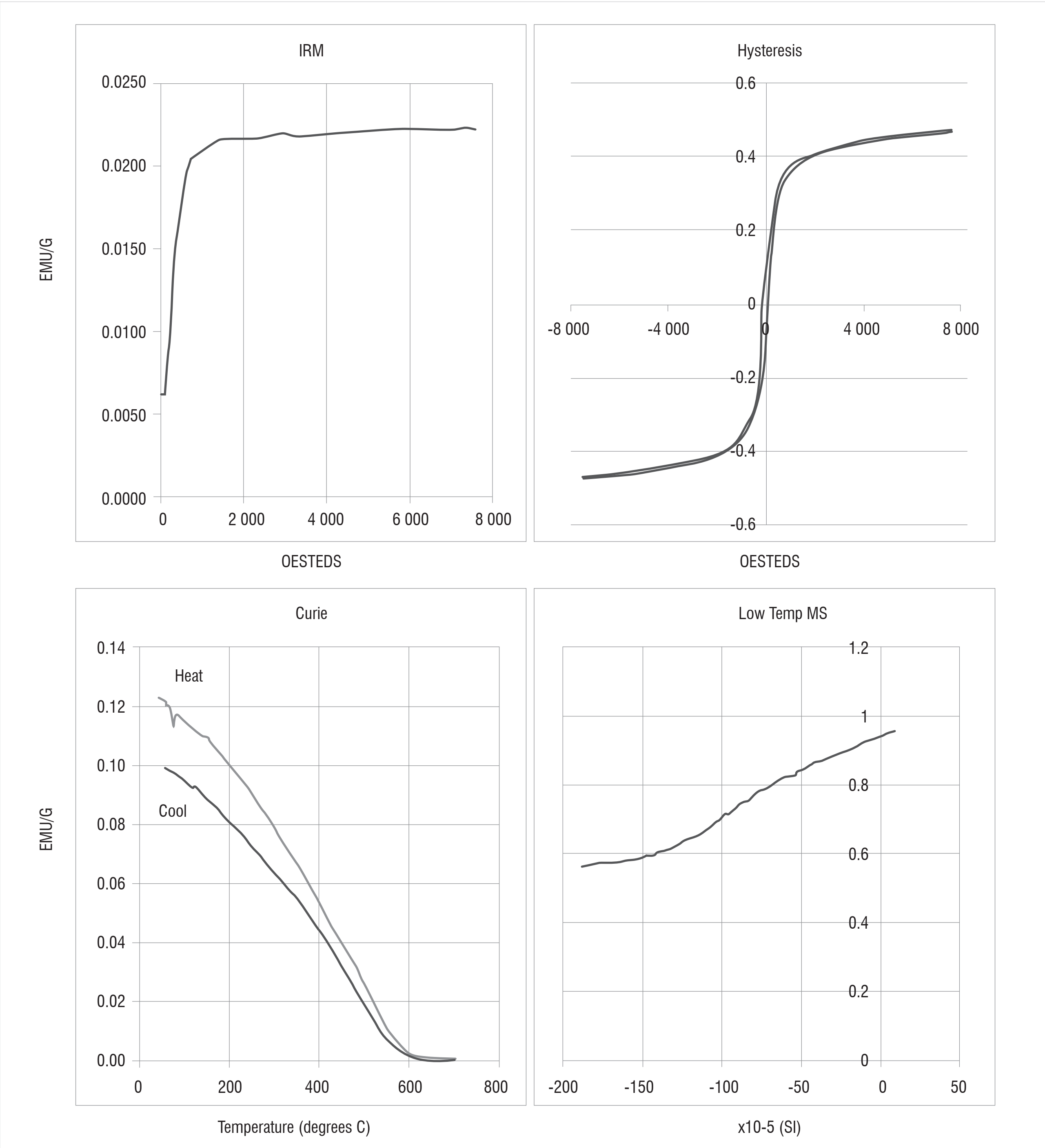
X-ray fluorescence microscopy (XFM) analyses were performed at the XFM beamline[45] at the Australian Synchrotron in order to investigate elemental distributions across speleothem cross sections. For the measurements, the beam size was set to ca. 5x5 mm^2 using Kirkpatrick-Baez mirrors. The incident photon energy was chosen as 18.5 keV, thus permitting excitation and detection of the trace metals of interest (Fe, Mn, As, Sr and others). Elemental maps were collected by scanning vertically mounted samples through the beam using x-y-translation stages while simultaneously collecting X-ray fluorescence from the sample with the 384-pixel Maia detector.[46] Maps were typically collected over cm^2 areas at a resolution of 5–10 mm, thus resulting in images of several megapixels size (for specific sampling details see corresponding figures).

Results

Mineral magnetic measurements (Figure 5) were undertaken on sister samples throughout the sequence to establish the main remanence carrier and possible effects caused by changes in mineralogy and magnetic grain size. The samples have extremely high magnetic susceptibility (χ_{LF}; up to 5.74 x10^{-6} m^3/kg^{-1}; Table 1) indicating significant proportions of ferromagnetic material, which is also confirmed by the high magnetic intensity of the remanence not only within the sediment samples, but also within the speleothems. The high magnetic susceptibility can be accounted for by the large proportion of fine- to ultra-fine-grained (<0.05 μm) single to superparamagnetic ferrimagnetic material as shown by frequency dependence of magnetic susceptibility (χ_{FD}%) values of between 6.20% and 15.8%, the latter of which is close to the maximum value expected for natural samples.[46] Isothermal remanent magnetisation acquisition curves and hysteresis loops further indicate the low coercivity, ferrimagnetic nature of the samples and Curie points of just less than 600 °C indicate the dominant mineral is magnetite.[48]

The presence of a low temperature tail below -150 °C in the low temperature magnetic susceptibility curves also indicates magnetite over maghaemite. The remanence within the samples is removed between 500 °C and 600 °C, which further confirms that it is primarily carried by magnetite. Low temperature magnetic susceptibility curves indicate no presence of a Verway transition and as such the low coercivity is interpreted to be as a result of viscous single domain grains close to the superparamagnetic boundary, as also indicated by high χ_{FD}%, rather than multi-domain grains. The stability of the samples during heating with little change in χ_{LF} and general reversibility of the thermomagnetic curves further supports this hypothesis, although a slight drop in magnetisation on cooling (Figure 5) in some samples may indicate the presence of maghaemite. A non-saturation of the IRM curves by 100 mT also indicates a proportion of larger stable single domain grains that carry the primary remanence.

X-ray fluorescence microscopy (XFM) studies at the Australian Synchrotron indicate that the majority of iron (Fe) phases occur as discrete horizontal bands within the flowstone (Figure 6) and correlate with higher concentrations of arsenic (Figure 7), which has been suggested to correlate to warmer, more humid phases with increased erosion.[49] This finding is further confirmed by the presence of manganese-bearing structures and their co-location with arsenic (Figure 7), the former of which would have come from greater dissolution of the dolomite host rock.



Scale: 2 mm x 2 mm

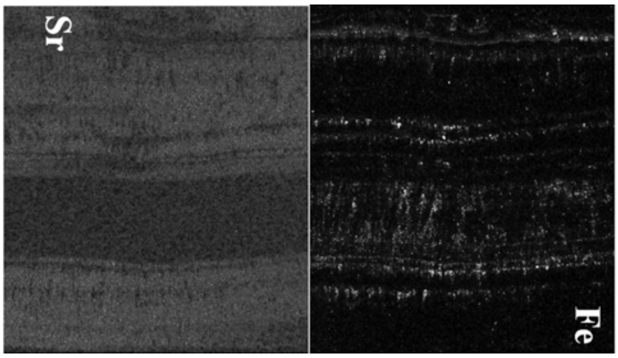
Figure 5: Mineral magnetic measurements – isothermal remanent magnetisation (IRM), hysteresis, thermomagnetic curve (Curie) and low temperature magnetic susceptibility (MS) – of a siltstone sample from Haasgat.

Table 1: Palaeomagnetic and mineral magnetic results from the Haasgat (HGT) fossil site

Sample	Section	Composite height (m)	χLF	χFD%	Declination	Inclination	Median angle of deviation	Number of samples	K	Palaeolatitude	Polarity	Age (Ma)	Deposit
HGTBWa	BSW	-0.30	2.26	11.62	18.9	-64.3	5.7	3	114	64.6	N	>1.78	Siltstone in breccia
HGTBWTb	BSW	-0.73	2.78	5.79	42.1	-63.7	2.6	3	78	51.0	IN		Siltstone in breccia
HGTBWc	BSW	-1.56	2.66	9.43	20.5	-44.9	5.7	3	205	71.5	N		Siltstone in breccia
HGTBEa	BSE	-2.08	3.44	9.24	329.0	59.1	13.5	3	4	18.5	I	~1.95	Siltstone in breccia
HGTBWd	BSW	-2.43	3.40	7.90	216.2	27.1	10.4	3	24	-54.3	IR		Siltstone in breccia
HGTBEb	BSE	-3.13	4.81	8.58	189.0	38.7	6.7	3	21	-81.0	R		Siltstone in breccia
HGTBWe	BSW	-3.73	1.39	5.32	208.5	52.1	5.3	5	36	-64.2	R		Flowstone
HGTLCa	LCS	-4.37	3.67	9.08	183.3	58.1	5.8	5	361	-76.4	R		Mudstone
HGTLCb	LCS	-4.72	4.67	8.58	176.9	63.7	2.4	3	238	-70.0	R		Mudstone
HGTTHa	THSS	-4.79	0.03	N/A	41.1	21.0	5.1	3	56.5	35.9	I		Flowstone
HGTTHb	THSS	-5.55	3.84	8.58	178.8	61.3	1.5	3	570	-73.0	R	<~2.3	Siltstone
HGTTHc	THSS	-5.90	2.44	7.62	175.8	61.1	4.4	4	30.4	-73.0	R		Siltstone
HGTTHe	THSS	-7.00	3.18	8.47	53.1	30.2	4.3	4	26.2	23.6	I		Siltstone

BSW, Bridge Section West; BSE, Bridge Section East; LCS, Lockwood's Crawl Section; THSS, Tetley's Hall Southern Section; N, normal; I, intermediate; R, reversed.

The XFM data equally shows that Fe-rich layers are strontium (Sr) poor and vice versa (Figure 6). This patterning indicates that the remanence of these bands of Fe is dominated by detrital material that was brought into the cave and deposited over the flowstones and then calcified into their growth structure. However, Fe is also scattered in a vertical pattern in some sections that appear to be related to crystal growth and as such some Fe may have been deposited as detrital material in the drip water or precipitated with the calcite. Weak Fe phases such as goethite can often be found holding a chemical remanence (CRM) within areas of calcite flowstone where no detrital contamination occurs[50,51] and this may be the case here. While such CRMs are generally masked where detrital contamination also occurs within the bands, such CRMs can account for low temperature components in thermal demagnetisation curves,[50-53] as goethite dehydrates to haematite on heating through 150-300 °C.



Scale: 1 mm x 6 mm

Figure 6: Synchrotron X-ray fluorescence microscopy data showing the negative correlation between iron (Fe) and strontium (Sr) in a flowstone at Haasgat. The Fe distribution (lighter in the figure) is dominated by the layering of the calcite in the flowstone and is likely detrital from flooding events between flowstone formation. However, it appears that some Fe also occurs along crystal growth boundaries between these laminated layers and may have been introduced into the flowstone during calcite formation. Areas of higher strontium are lighter.

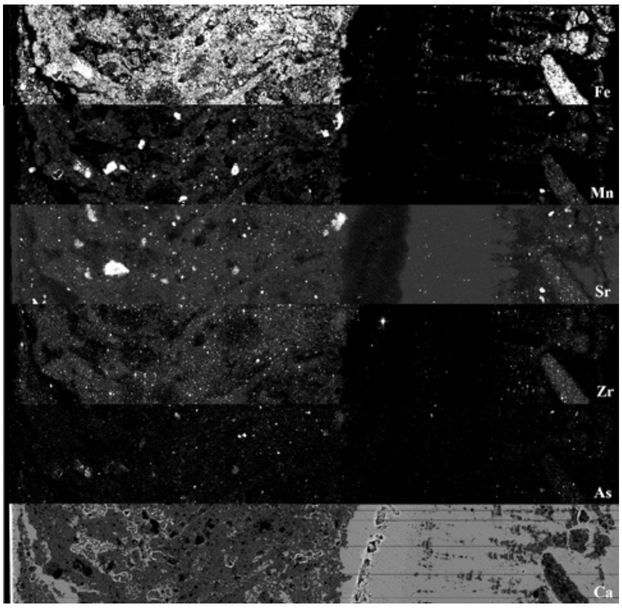


Figure 7: Synchrotron X-ray fluorescence data showing the correlation of iron (Fe) with manganese (Mn) and arsenic (As) versus the structure of the calcite (Ca; calcium) in the flowstone. The high Fe portion on the left of the specimen represents a thick flood layer within the flowstone. In the right-hand side of the specimen, large Fe-rich sediment clasts can be seen occurring within the Ca-rich portion of the flowstone. In the lower section, Fe can also be seen occurring as another thinner layer within the flowstone, as well as along crystal boundaries. (Note: the darker the image the less there is of each element.)

Fe is quite regularly distributed within the Haasgat flowstones (Figure 6) and as such the remanence of standard 25-mm palaeomagnetic cores is measuring the remanence of multiple flood events. In contrast to the calcified sediments which hold a post-depositional remanent magnetisation formed after dewatering and compaction of the sediments, the remanence in speleothem is a true depositional remanent magnetisation that is formed extremely soon after precipitation over and calcification of the flooding sediment surface, perhaps within a few days or weeks.[50,52,53] There is also some evidence in the XFM data that Fe may have been deposited in the flowstones from precipitation from water as Fe seems to also be dispersed within the calcite crystal growth patterns (Figure 5). This material is likely to be instantaneously locked in and should provide an extremely good record of the ancient field and can often be more reliable than the sediments that surround them because mud cracks, slope effects and other features can cause inclination shallowing or changes in the declination values during post-depositional compaction. The upper flowstone sampled at Haasgat from the Bridge Section records a stable reversed polarity direction (Table 1). However, the flowstone sampled from Tetley's Hall records an intermediate polarity. Based on the XFM data, there is no reason to suggest that this result is because of the recording medium, and it is possible that this result represents a true field deviation and perhaps the edge of a short reversal event or excursion.

The majority of deposits from the site formed during a long period of reversed magnetic polarity (Table 1; Figures 4 and 8). The majority of samples record a stable magnetic polarity with low median angle of deviation values that is both consistent between different sub-samples from the same block (as shown by the low K values) and between samples from different blocks in the same part of the sequence. Reversed samples have declinations of 176–216° and inclinations between +50° and +64°. Normal polarity samples have declinations between 18° and 45° and inclinations of -45° to -69° and so the results show good dipolar dispersal. Some samples do have much shallower mean inclination values (-27° to -45°) that could be a result of post-depositional effects; however, these samples are all on either side of the reversal and they are more likely true field directions related to the polarity reversal. Some individual samples did show inclination shallowing that is more likely related to wetting and drying cycles and fossil mud-cracks are seen within the layers. Directions that are entirely intermediate in nature are also recorded, notably during the reversal from reversed to normal polarity in the Bridge Section, and this definitely represents geomagnetic field change. For a very high proportion of samples (>90%), a stable remanence was recorded (Table 1; Figure 7), which is in part a result of the well-laminated siltstone and flowstone nature of the majority of the deposits and the fact that more brecciated layers, notably in the Bridge, were avoided during sampling. However, the large amount of breccia has led to larger gaps in this part of the sequence. Samples from the Bridge Sections record reversed polarities at the base and normal polarities at the top (Table 1). Samples from the exposures within Lockwood's Crawl (LCS) all record reversed polarity as do sediments containing dense layers of fossils in the top of THSS. Below the fossil layers, the deposits contain a mixture of reversed and intermediate polarities. It is noticeable that well-indurated samples from Tobias' Steps record reversed polarities while those from the same layers exposed in THSS give intermediate polarities. This finding appears to be a result of decalcification and recrystallisation of the calcite within the THSS deposits related to the later phase of passage formation outlined above. In sum, the Haasgat sequence indicates a change from a longer period of reversed polarity in the lower Eastern Deposits, punctuated by periods of intermediate polarity towards its base, to normal polarity in the upper Western Deposits in the Bridge.

Discussion and conclusions

With two species of *Equus* recovered within the ex-situ HGD faunal sample[22], and a nearly complete *Equus* metacarpal exposed in the in-situ calcified sediment sequence at a level equivalent to reversed polarity layers in THSS, the majority of the fossiliferous deposits are likely not older than ~2.3 Ma (the oldest occurrence of *Equus* in Africa[30,31]) and the majority of other fauna suggest the site is older than Gondolin GD2 at 1.95–1.78 Ma[1,22,35]. An age assessment for the deposits based on magnetobiostratigraphy requires the identification of a longer period of reversed polarity as seen in the Western Deposits and the base of the Bridge Deposits followed by a reversal to a period of normal polarity as seen in the top of the Bridge Deposits (Figures 4 and 9). Only one major reversal from reversed to normal polarity occurs within the 2.3–1.8 Ma time period, at the beginning of the Olduvai SubChron at ~1.95 Ma[54,55] (Figure 9). While several short geomagnetic field events also occur in this time period at ~1.98 Ma (Pre-Olduvai), ~2.05–2.02 Ma (Huckleberry Ridge) and ~2.21–2.14 Ma (Réunion), and have been found in other South African palaeocaves,[3,6,10] they almost exclusively occur in speleothem, not cave sediments. The same magnetic reversal has been found in flowstone speleothem in different caves, suggesting that such flowstones could be used as marker beds between sites, akin to the volcanic tuffs of East Africa. As a result of the potential rapid accumulation of sediments, compared to flowstone, the normal polarity period (which covers about 2 m of the total 7 m of section analysed) could represent such short polarity episodes (generally <6 ka in length[54]), but it is much less likely. The chances of the sediments forming during such a short reversal are highly unlikely when compared to the more parsimonious explanation that they were deposited during a longer period (~170 ka) of stable polarity such as the Olduvai SubChron (1.95–1.78 Ma). That being said, such a short period has been suggested to encompass the deposition of the Malapa fossils[10], even though deposition within the Olduvai SubChron was previously considered more likely[9]. The intermediate polarities noted in speleothems towards the base of the sequence (in THSS) are more likely to relate to such short reversal events that were perhaps not entirely recorded or have been missed during the current sampling phase at the site. Given the short lock in time of speleothem and the fact that the XFM data indicate that Fe distribution in the speleothems is within discrete bands, the intermediate polarities (Figure 8c) are considered to be true field directions rather than an effect of the recording medium.

Further detailed sampling will concentrate on increasing the sample resolution in this area of the cave. As such, the Eastern Bridge Deposits at the front of the cave most likley date to either side of 1.95 Ma, although an older age between 2.21 Ma and 1.98 Ma cannot be ruled out at this stage. The fossil rich Western Deposits (EAS, LCS, THSS) at the rear of the cave all date to a period prior to 1.95 Ma. The question that remains is how much older the Western Deposits are than 1.95 Ma, given that the next oldest major reversal is the Gauss–Matuyama boundary at 2.58 Ma. The occurrence of *Equus* would suggest that it is unlikely to be older than ~2.3 Ma, although in-situ *Equus* remains have not been recovered from the base of the sequence deeper than ~5 m. More detailed palaeomagnetic analysis could potentially identify one or more of the small geomagnetic reversals in this time period (~2.16 Ma, ~2.04 Ma, ~1.98 Ma[3,6,10]) but this identification will require extremely detailed sampling of the deposits. Uranium-lead dating of flowstones in the base of the sequence would also help refine the age.

Given the occurrence of the reversal at the beginning of the Olduvai SubChron at ~1.95 Ma (or other reversal between 2.21 Ma and 1.98 Ma), it is also possible that the whole deposit and its fossils may not be much older than this reversal, especially given that it is currently unclear exactly how the Western and Eastern Deposits correlate exactly as the stratigraphy is partly obscured between them. The Western Deposits are certainly older than the reversal identified but they cannot be directly stacked beneath the Eastern Deposits to create a composite stratigraphy as the siltstones of the Western Deposits were in part formed by winnowing of the Eastern breccias and so are partly contemporary. It is possible that flowstone layers towards the base of the Eastern Deposits correlate with the flowstone deposition also noted in the middle of the Western Sequence. If this were the case then the reversed polarity period would go from covering ~5 m of stratigraphy to ~4 m, although this is still double that covered by the normal polarity period. Further detailed sampling and stratigraphic work will hopefully resolve many of these issues but at present a good estimate for the age of the deposit is between 2.3 Ma and 1.8 Ma, perhaps around ~2 Ma.

At an age between 2.3 Ma and 1.8 Ma, Haasgat is potentially pene-contemporaneous with a number of other deposits and fossils in South Africa (Figure 9) including *Australopithecus sediba* at Malapa at ~1.98 Ma[10], the last occurrence of *Australopithecus africanus* (Sts 5; 'Mrs Ples') in Sterkfontein Member 4 at ~2.05–2.02 Ma[3,6] and the first occurrence of early *Homo* (SK 847), *Paranthropus robustus* and the Oldowan tool industry in the Hanging Remnant of Swartkrans Member 1

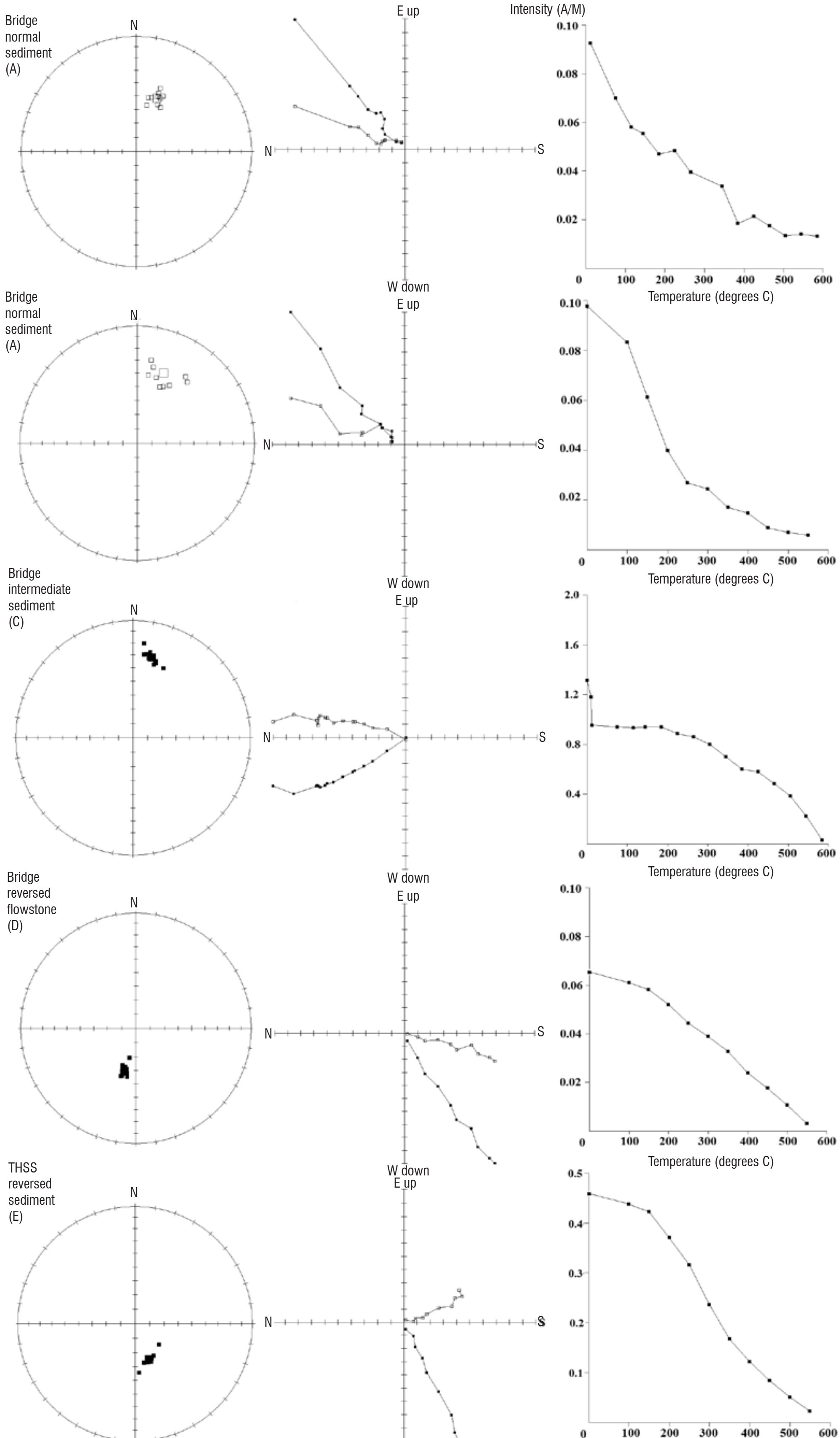


Figure 8: Demagnetisation spectra (stereo, Zijderveld and intensity plots) for representative (a,b) normal, (c) intermediate and (d,e) reversed polarity siltstone within (a–c) breccia, (e) mudstone and (d) flowstone samples from Haasgat.

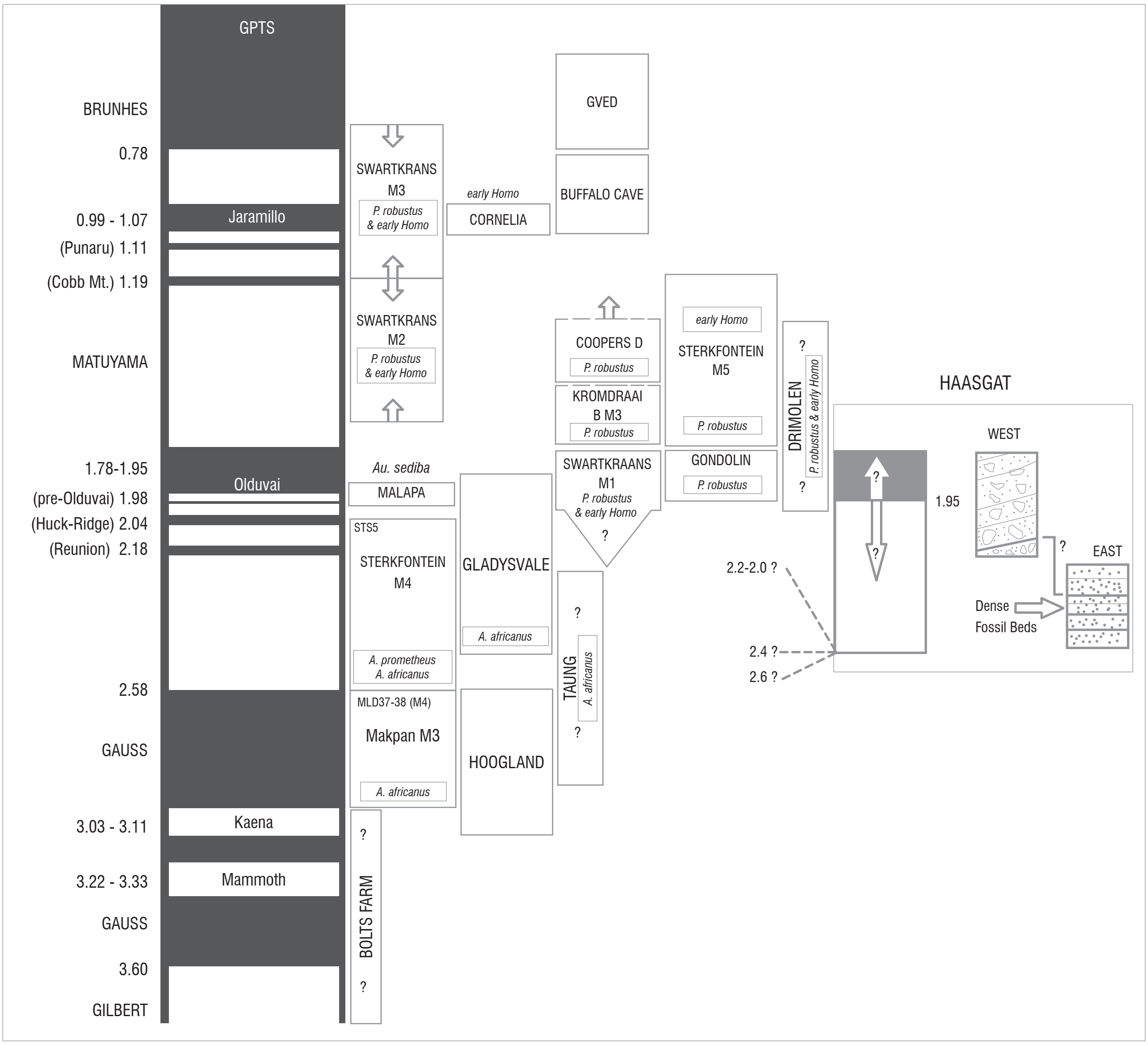


Figure 9: The age of Haasgat on the geomagnetic polarity timescale (GPTS) compared with other hominin and vertebrate fossil bearing sites.

sometime between 2.25 Ma and 1.80 Ma, although more likely between 2.0 Ma and 1.8 Ma[1,7]. Haasgat may also be contemporary with the oldest deposits at the nearby Gladysvale Cave site which are as old as ~2.4 Ma[32] and Member 1 at Kromdraai B, which also records the beginning of the Olduvai SubChron at ~1.95 Ma[56]. While it has been suggested that the ex-situ recovered TM1517 type specimen of *Paranthropus robustus* derives from Kromdraai B Member 1 based on the colour of the adhering matrix[56], all the in-situ hominins from Kromdraai B have come from Member 3, which is estimated to be younger than 1.8 Ma[1]. The Haasgat deposits, with a unique representation of primate and ungulate species, are thus dated to an extremely significant period for human evolution in South Africa with the occurrence of a major turnover in hominin species (*Australopithecus* to *Paranthropus* and *Homo*) and the first occurrence of stone tools that may be related to major climatic and environmental shifts and increased aridity since ~2.2 Ma.[57]

Acknowledgements

This paper and the naming of different portions of the Haasgat site are dedicated to the memory of our recently departed colleagues Andre Keyser (who first worked the Haasgat site), Tim Partridge and Philip Tobias, who all made significant contributions towards understanding the geology and fossil record of South Africa, as well as our colleagues Elizabeth Harmon and Charles Lockwood, whose contributions in the field of hominin anatomy and untimely loss will continue to impact all of us for decades to come. Excavation, survey and fossil sampling and preparation at Haasgat were additionally supported by Lazarus Kgasi of the Plio-Pleistocene Section, Department of Vertebrates, Ditsong National Museum of Natural History in Pretoria under SAHRA permit 80/10/03/010/51 held by Stephany Potze. This project was funded by the Australian Research Council Future Fellowship Grant FT120100399 to A.I.R.H. and a National Science Foundation Grant (NSF BCS 0962564) to J.W.A. The synchrotron analysis was undertaken on the XFM beamline at the Australian Synchrotron (beamtime granted to A.I.R.H. and P.K.). The work was supported by Mimi Hill at the University of Liverpool Geomagnetism Laboratory. We are indebted to Phil Tetley, his family, and the entire Kalkheuvel West community for their support and hospitality during our continued work at Haasgat and SAHRA for issuing us a permit to work at the site. We thank three anonymous reviewers for detailed comments.

Authors' contributions

A.I.R.H. undertook the palaeomagnetic analysis and geological sampling of the site, aided in the latter by J.W.A. J.W.A. and A.D.T.K. undertook the faunal analysis and biochronology of material from the site. A.I.R.H., P.K., D.P., D.L.H. and M.D.d.J. undertook the synchrotron XFM work on speleothem from the site. J.W.A., A.I.R.H. and S.P. led the team, decided on site naming and descriptions and recovered the fossils from the site; S.P. holds the permit for the site. S.P. also undertook the acid preparation of the fossil material.

References

1. Herries AIR, Adams JW. Clarifying the context, dating and age range of the Gondolin hominins and *Paranthropus* in South Africa. J Hum Evol. 2013;65:676–681. http://dx.doi.org/10.1016/j.jhevol.2013.06.007

2. Herries AIR, Curnoe D, Adams JW. A multi-disciplinary seriation of early *Homo* and *Paranthropus* bearing palaeocaves in southern Africa. Quat Int. 2009;202:14–28. http://dx.doi.org/10.1016/j.quaint.2008.05.017

3. Herries AIR, Hopley P, Adams JW, Curnoe D, Maslin M. Geochronology and palaeoenvironments of the South African early hominin bearing sites: A reply to 'Wrangham et al., 2009: Shallow-water habitats as sources of fallback foods for hominins'. Am J Phys Anthropol. 2010;143:640–646. http://dx.doi.org/10.1002/ajpa.21389

4. Clarke RJ. Latest information on Sterkfontein's *Australopithecus* skeleton and a new look at *Australopithecus.* S Afr J Sci. 2008;104:443–449.

5. Pickering R, Kramers J. Re-appraisal of the stratigraphy and determination of new U-Pb dates for the Sterkfontein hominin site, South Africa. J Hum Evol. 2010;59:70–86. http://dx.doi.org/10.1016/j.jhevol.2010.09.001

6. Herries AIR, Shaw J. Palaeomagnetic analysis of the Sterkfontein palaeocave deposits: Age implications for the hominin fossils and stone tool industries. J Hum Evol. 2011;60:523–539. http://dx.doi.org/10.1016/j.jhevol.2010.09.001

7. Pickering R, Kramers JD, Hancox PJ, de Ruiter DJ, Woodhead JD. Contemporary flowstone development links early hominin bearing cave deposits in South Africa. Earth Planet Sci Lett. 2011;306:23–32. http://dx.doi.org/10.1016/j.epsl.2011.03.019

8. Berger LR, de Ruiter DJ, Churchill SE, Schmid P, Carlson KJ, Dirks PHGM, et al. *Australopithecus sediba*: A new species of *Homo*-like australopith from South Africa. Science. 2010;328:195–204. http://dx.doi.org/10.1126/science.1184944

9. Dirks PHDM, Kibii JN, Kuhn BF, Steininger C, Churchill SE, Kramers JD, et al. Geological setting and age of *Australopithecus sediba* from southern Africa. Science. 2010;328:205–208. http://dx.doi.org/10.1126/science.1184950

10. Pickering R, Dirks P, Jinnah Z, de Ruiter DJ, Churchill SE, Herries AIR, et al. *Australopithecus sediba* at 1.977 Ma and implications for the origins of the genus *Homo*. Science. 2011;333:1421–1423.

11. Hartstone-Rose A, Kuhn BF, Nalla S, Werdelin L, Berger LR. A new species of fox from the *Australopithecus sediba* type locality, Malapa, South Africa. Trans R Soc S Afr. 2013;68(1):1–9. http://dx.doi.org/10.1080/0035919X.2012.748698

12. Gommery D, Thackeray JF, Sénégas F, Potze S, Kgasi L. The earliest primate (*Parapapio* sp.) from the Cradle of Humankind World Heritage site (Waypoint 160, Bolt's Farm, South Africa). S Afr J Sci. 2008;104(9–10):405–408.

13. Gommery D, Thackeray JF, Sénégas F, Potze S, Kgasi L. Additional fossils of *Parapapio* sp. Waypoint From 160 [Bolt's Farm, South Africa], dated between 4 and 4.5 million years ago. Ann Transv Mus. 2009;46:63–72.

14. Herries AIR, Adams JW, Kuykendall KL, Shaw J. Speleology and magnetobiostratigraphic chronology of the GD 2 locality of the Gondolin hominin-bearing palaeocave deposits, North West Province, South Africa. J Hum Evol. 2006;51:617–631. http://dx.doi.org/10.1016/j.jhevol.2006.07.007

15. Adams JW, Herries AIR, Conroy GC, Kuykendall KL. Taphonomy of a South African cave: Geological and hydrological influences on the GD 1 fossil assemblage at Gondolin, a Plio-Pleistocene palaeocave system in the Northwest Province, South Africa. Quat Sci Rev. 2007;26:2526–2543. http://dx.doi.org/10.1016/j.quascirev.2007.05.006

16. Adams JW. Taphonomy of the Gondolin GD 2 in-situ deposits and its bearing on interpretations of South African Plio-Pleistocene karstic fossil assemblages. J Taphonomy. 2010;8:81–116.

17. Adams JW, Herries AIR, Hemingway J, Kegley A, Kgasi, L, Hopley P, et al. Initial fossil discoveries from Hoogland, a new Pliocene primate-bearing karstic system in Gauteng Province, South Africa. J Hum Evol. 2010;59:685–691. http://dx.doi.org/10.1016/j.jhevol.2010.07.021

18. Brock A, McFadden PL, Partridge TC. Preliminary palacomagnetic results from Makapansgat and Swartkrans. Nature. 1977;266:249–250. http://dx.doi.org/10.1038/266249a0

19. McFadden PL, Brock A, Partridge TC. Palaeomagnetism and the age of the Makapansgat hominid site. Earth Planet Sci Lett. 1979;44:373–382. http://dx.doi.org/10.1016/0012-821X(79)90076-1

20. Herries AIR, Reed K, Kuykendall KL, Latham AG. Speleology and magnetobiostratigraphic chronology of the Buffalo Cave fossil bearing palaeodeposits, Makapansgat, South Africa. Quat Res. 2006;66:233–245. http://dx.doi.org/10.1016/j.yqres.2006.03.006

21. Lacruz RS, Brink JS, Hancox J, Skinner AS, Herries A, Schmidt P, et al. Palaeontology, geological context and palaeoenvironmental implications of a Middle Pleistocene faunal assemblage from the Gladysvale Cave, South Africa. Palaeontol Afr. 2002;38:99–114.

22. Adams JW. A revised listing of fossil mammals from the Haasgat cave system ex situ deposits (HGD), South Africa. Palaeontol Electr. 2012;15(3), Article 29A, 88 pages.

23. Keyser AW, Martini JEJ. Haasgat, a new Plio-Pleistocene fossil locality. Paleoec Aft. 1991;21:119–129.

24. Keyser AW. The palaeontology of Haasgat: A preliminary account. Palaeontol Afr. 1991;28:29–33.

25. McKee J, Keyser AW. Craniodental remains of *Papio angusticeps* from the Haasgat cave site, South Africa. Int J Primatol. 1994;15:823–841. http://dx.doi.org/10.1016/j.jhevol.2010.08.002

26. Plug I, Keyser AW. Haasgat Cave, a Pleistocene site in the central Transvaal: geomorphological, faunal and taphonomic considerations. Ann Trans Mus. 1994;36:139–145.

27. Jablonski N, Frost S. Cercopithecoidea. In: Werdelin L, Sanders W, editors. Cenozoic mammals of Africa. Berkeley: University of California Press; 2010. pp. 393–428. http://dx.doi.org/10.1525/california/9780520257214.003.0023

28. McKee J, Von Mayer A, Kuykendall K. New species of *Cercopithecoides* from Haasgat, North West Province, South Africa. J Hum Evol. 2011;60:83–93.

29. de Ruiter D, Pickering R, Steininger C, Kramers J, Hancox P, Churchill S, et al. New *Australopithecus robustus* fossils and associated U-Pb dates from Cooper's Cave (Gauteng, South Africa). J Hum Evol. 2009;56:497–513. http://dx.doi.org/10.1016/j.jhevol.2009.01.009

30. Geraads D, Raynal J, Eisenmann V. The earliest human occupation of North Africa: A reply to Sahnouni et al. (2002). J Hum Evol. 2004;46:751–761. http://dx.doi.org/10.1016/j.jhevol.2004.01.008

31. Brown FH, Mcdougall I. Geochronology of the Turkana Depression of Northern Kenya and Southern Ethiopia. Evol Anthropol. 2011;20:217–227. http://dx.doi.org/10.1002/evan.20318

32. Herries AIR, Pickering R, Adams JW, Curnoe D, Warr G, Latham AG, et al. A multi-disciplinary perspective on the age of *Australopithecus* in southern Africa. In: Reed KE, Fleagle JG, Leakey R, editors. Paleobiology of *Australopithecus*: Contributions from the Fourth Stony Brook Human Evolution Symposium and Workshop, Diversity in *Australopithecus*: Tracking the First Bipeds. Vertebrate Paleobiology and Paleoanthropology series. Dordrecht: Springer; 2013. p. 21–40.

33. Harris J, White T. Evolution of the Plio-Pleistocene African Suidae. Trans Am Phil Soc. 1979;69:5–128. http://dx.doi.org/10.2307/1006288

34. Brink J, Herries AIR, Moggi-Cecchi J, Gowlett J, Adams JW, Hancox J, et al. First hominine remains from ~1 Ma bone bed at Cornelia-Uitzoek, Free State Province, South Africa. J Hum Evol. 2012;63:527–535. http://dx.doi.org/10.1016/j.jhevol.2012.06.004

35. Adams JW. A methodology for the intraspecific assessment of heterogeneously worn hypsodont teeth using computerized tomography. J Taphonomy. 2005;3:151–612.

36. Kegley ADT, Hemingway J, Adams JW. Odontometric analysis of the reanalyzed and expanded *Cercopithecoides* from the Haasgat fossil assemblage, Cradle of Humankind, South Africa. Am J Phys Anthropol. 2011;S144:183.

37. Latham AG, Herries AIR, Kuykendall K. The formation and sedimentary infilling of the Limeworks Cave, Makapansgat, South Africa. Palaeontol Afr. 2003;39:69–82.

38. Latham AG, Herries A, Quinney P, Sinclair A, Kuykendall K. The Makapansgat australopithecine site from a speleological perspective. In: Pollard AM, editor. Geoarchaeology: Exploration, environments, resources. London: Royal Geological Society, Special Publications. 1999;165:61–77.

39. Latham AG, Herries AIR. The formation and sedimentary infilling of the Cave of Hearths and Historic Cave Complex. Geoarchaeology. 2004;19:323–342. http://dx.doi.org/10.1002/gea.10122

40. Herries AIR. A chronological perspective on the Acheulian and its transition to the Middle Stone Age in southern Africa: the question of the Fauresmith. Int J Evol Biol. 2011;961401, 25 pages.

41. Braun D, Levin N, Stynder D, Herries AIR, Archer W, Forrest F, et al. Mid-Pleistocene hominin occupation of Elandsfontein, Western Cape of South Africa. Quat Sci Rev. 2013;82:145–166. http://dx.doi.org/10.1016/j.quascirev.2013.09.027

42. Vrba ES. The fossil record of African antelopes (Mammalia, Bovidae) in relation to human evolution and paleoclimate. In: Vrba ES, Denton GH, Partridge TC, Burckle LH, editors. Paleoclimate and evolution with emphasis on human origins. New Haven: Yale University Press; 1995. p. 385–424.

43. Herries AIR, Latham AG, Kuykendall K. The use of 'SRT' in palaeomagnetic sampling of the Makapansgat limeworks hominid palaeocave, South Africa. Antiquity. 2001;75:251–252.

44. International Geomagnetic Reference Field: The eleventh generation. International Association of Geomagnetism and Aeronomy, Working Group V-MOD. Participating members: C. C. Finlay, S. Maus, C. D. Beggan, T. N. Bondar, A. Chambodut, T. A. Chernova, A. Chulliat, V. P. Golovkov, B. Hamilton, M. Hamoudi, R. Holme, G. Hulot, W. Kuang, B. Langlais, V. Lesur, F. J. Lowes, H. Luhr, S. Macmillan, M. Mandea, S. McLean, C. Manoj, M. Menvielle, I. Michaelis, N. Olsen, J. Rauberg, M. Rother, T. J. Sabaka, A. Tangborn, L. Toffner-Clausen, E. Thebault, A. W. P. Thomson, I. Wardinski Z. Wei, T. I. Zvereva. Geophys J Int. 2010;183(3):1216–1230. DOI: 10.1111/j.1365-246X.2010.04804.x.

45. Paterson D, De Jonge MD, Howard DL, Lewis W, McKinlay J, Starritt A, et al. The X-ray fluorescence microscopy beamline at the Australian synchrotron. AIP Conf Proc. 2011;1365:219–222. http://dx.doi.org/10.1063/1.3625343

46. Kirkham R, Dunn PA, Kuczewski A, Siddons DP, Dodanwela R, Moorhead G, et al. The Maia spectroscopy detector system: Engineering for integrated pulse capture, low-latency scanning and real-time processing. AIP Conf Proc. 2010;1234:240–243. http://dx.doi.org/10.1063/1.3463181

47. Walden J, Oldfield F, Smith J. Environmental Magnetism: A practical guide. Technical Guide 6. London: Quaternary Research Association; 1999.

48. Dunlop DJ, Özdemir Ö. Rock magnetism: Fundamentals and frontiers. Cambridge: Cambridge University Press; 2001.

49. Zhou H, Greig A, You C-F, Lai Z, Tang J, Guan Y, et al. Arsenic in a speleothem from central China: Stadial-interstadial variations and implications. Environ Sci Technol. 2011;45:1278–1283. http://dx.doi.org/10.1021/es1032103

50. Lascu I, Feinberg JM. Speleothem magnetism. Quat Sci Rev. 2011;30:3306–3320. http://dx.doi.org/10.1016/j.quascirev.2011.08.004

51. Pickering R, Jacobs Z, Herries AIR, Karkanas P, Bar-Matthews M, Woodhead JD, et al. Paleoanthropologically significant South African sea caves dated to 1.1-1.0 million years using a combination of U-Pb, TT-OSL and palaeomagnetism. Quat Sci Rev. 2013;65:39–52. http://dx.doi.org/10.1016/j.quascirev.2012.12.016

52. Latham AG. Paleomagnetism, rock magnetism and U-Th dating of speleothem deposits [PhD thesis]. Ontario: McMaster University; 1981.

53. Latham AG, Ford D. The palaeomagnetism and rock magnetism of cave and karst deposits. In: Aissaoui DM, McNeill DF, Hurley NF, editors. Applications of paleomagnetism to sedimentary geology. Tulsa, OK: Society of Economic Paleontologists and Mineralogists; 1993. p. 150–155.

54. Laj C, Channell JET. Geomagnetic reversals. In: Kono M, editor. Treatise on geophysics: Geomagnetism. Amsterdam: Elsevier; 2007. pp. 373–416. http://dx.doi.org/10.1016/B978-044452748-6.00095-X

55. Gradstein FM, Ogg JG, Schmitz M, Ogg G. The geologic time scale. Boston: Elsevier; 2012.

56. Thackeray JF, Kirschvink JL, Raub TD. Palaeomagnetic analysis of calcified deposits from the Plio-Pleistocene hominid site of Kromdraai, South Africa. S Afr J Sci. 2002;98:537–540.

57. Dupont LM, Donner B, Vidal L, Pérez EM, Wefer G. Linking desert evolution and coastal upwelling: Pliocene climate change in Namibia. Geology. 2005;33:461–464. http://dx.doi.org/10.1130/G21401.1

Engine emissions and combustion analysis of biodiesel from East African countries

AUTHOR:
Paul Maina[1,2]

AFFILIATIONS:
[1]Department of Mechanical Engineering, Tshwane University of Technology, Pretoria, South Africa
[2]Department of Mechanical and Production Engineering, Moi University, Eldoret, Kenya

CORRESPONDENCE TO:
Paul Maina

EMAIL:
mainap@tut.ac.za

POSTAL ADDRESS:
Department of Mechanical Engineering, Tshwane University of Technology, Private Bag X680, Pretoria 0001, South Africa

DATES:
Received: 06 Apr. 2013
Revised: 15 Oct. 2013
Accepted: 20 Oct. 2013

KEYWORDS:
biodiesel; combustion characteristics; emissions; fuel; *Jatropha*

HOW TO CITE:
Maina P. Engine emissions and combustion analysis of biodiesel from East African countries. S Afr J Sci. 2014;110(3/4), Art. #2013-0097, 8 pages. http://dx.doi.org/10.1590/sajs.2014/20130097

Environmental, availability and financial problems associated with fossil fuels encourage the manufacture and use of biodiesel. In this study, vegetable oil was extracted from *Jatropha curcas* seeds sourced from Kenya and Tanzania. A two-step acid–base catalytic transesterification process was used to produce biodiesel because of the amount of free fatty acids present in the oil. The test rig used in the experiments was an Audi, 1.9-litre, turbocharged direct injection, compression ignition engine. Emissions were measured using an Horiba emission analyser system while combustion data was collected by a data acquisition system, from which cylinder pressure and rate of heat release of the test engine in every crank angle were calculated. The two biodiesels showed better emission characteristics than the fossil diesel included in the tests for comparison purposes. Cylinder pressure and heat release of the biodiesel were also within acceptable ranges. However, the emission and combustion characteristics differed between the two biodiesels – a result likely related to their different origins. These findings prove that the source of biodiesel is an important factor to consider.

Introduction

Fuel combustion

Any material that can be burned to release thermal energy is called a fuel. Most familiar fuels consist primarily of hydrogen and carbon, and as such are called hydrocarbon fuels and exist in all phases. Combustion is a chemical reaction during which a fuel is oxidised and a large quantity of energy is released. The combustion process takes place in a controlled manner. The oxidiser most often used in combustion processes is air because it is free and readily available. Pure oxygen (O_2) is used as an oxidiser only in some specialised applications, such as cutting and welding, for which air cannot be used. A fuel must be brought above its ignition temperature to start combustion and the proportions of the fuel and air must be in the proper range for combustion to be initiated.

All combustible components of a fuel are burned to completion during a complete combustion process. The combustion process depends on many different parameters, such as the volume of injected fuel, the timing of ignition delay and air charge mixtures. In diesel engines, the timing of fuel injection is a major parameter that affects the combustion and exhaust emissions. If the injection of fuel begins earlier, the initial air temperature and pressure are lower; as a result, the ignition delay will be longer. If the injection starts later (when the piston is closer to the top dead centre), the initial air temperature and pressure are higher and, consequently, the ignition delay is shorter. Hence, the variation in injection timing exerts a major influence on the engine performance and exhaust emissions. Emissions from the combustion of these fuels are carbon dioxide (CO_2), carbon monoxide (CO), unburnt hydrocarbons, nitrogen oxides (NO_x), particulate matter, sulphur dioxide (SO_2) and other toxic compounds. The amount and type of emissions attributable to fuel combustion depend very much on the type of fuel burned.

Why biodiesel

Environmental protection laws continue to be more and more stringent, thus making the use of fossil fuels less favourable. These laws, in addition to the scarcity of fossil fuel which increases its cost, have prompted researchers and industries to search for new and better fuels. Biodiesel, a mixture of mono alkyl esters of long-chain fatty acids, derived from renewable lipid feedstock, such as vegetable oil or animal fat, is one of the alternative fuels which has received a lot of attention because it can be used in diesel engines without major modifications. It offers many advantages in that it is renewable, energy efficient, non-toxic, sulphur free and biodegradable and it burns more cleanly. Indirectly, by encouraging the planting of trees which consume the CO_2 produced during combustion, the use of biodiesel also reduces global warming gas emissions. Biodiesel as a fuel is important, especially in the rural areas of agriculturally developing countries like Kenya and Tanzania where there is vast unutilised land and a lack of modern forms of energy. Information on the production and characterisation of biodiesel from these countries is sparse.

Factors influencing biodiesel properties

It is known that the type of biodiesel produced from vegetable oil depends on the type of vegetable oil and its geographical location. For instance, several researchers have found that jatropha oil produces a superior biodiesel to that of many other non-edible feedstocks.[1-3] Many researchers also have observed different characteristics in biodiesels produced in the same way from similar plants from different regions. For example, Emil et al.[4] compared jatropha seed oil from Malaysia, Indonesia and Thailand which was produced in the same way using the same facilities, and found that the physicochemical characteristics of the oil produced were very different; the free fatty acid content, iodine value, saponification value, oil content and density were different in oils from different regions. Lu et al.[5] produced biodiesel from jatropha oil originating in China and observed that the properties of the crude jatropha oils varied with their region of origin in China. They stated that, because fatty acid content affects the biodiesel production process and the biodiesel fuel properties, jatropha oils from different origins which have different fatty acid contents, leads to biodiesels with different properties.

In the same context, Foidl et al.[6] produced jatropha oil from seeds which hailed from Cape Verde and Nicaragua. The seeds were processed in a similar manner. The oils were produced using the same equipment and in a similar environment. Still, the properties and content characteristics of the produced oils were different. The difference was conspicuous, ranging from the composition of the dried seeds to the saponification number, viscosity, free fatty acid content and iodine number. Vegetable oils other than jatropha oil have shown similar results: Lalas and Tsaknis[7] compared the characteristics of oils from *Moringa oleifera* of Indian origin and Kenyan (Mbololo) origin and found that, even though the oil extraction and processing procedure were identical, the properties of the oils differed according to the regions from which they were sourced. In addition to the *Jatropha* tree having excellent adaptation capacity, jatropha oil produces superior biodiesel compared with those of many other non-edible feedstocks, which is why it was used in this study.

Aim

Plants from different regions produce oils with characteristics which vary by wide margins and depend more on the regions in which the plants are grown than the species from which the oils are derived, most probably as a result of the environmental conditions to which they are exposed. It is thus very important to study the potential, characteristics and operation of biodiesel produced from a plant (e.g. *Jatropha*) originating from a different region. In this study, two biodiesels were produced – in the same lab using the same chemicals and equipment – from jatropha oil from two neighbouring East African countries. The characteristics of the two biodiesels were compared in terms of emissions and combustion. To understand and optimise the combustion process, a careful analysis of cylinder pressure and heat release was performed, which furnished precise information about the combustion process.

Materials and equipment

Materials

Samples of *Jatropha curcas* oil were supplied from Kenya (Jatropha A) and Tanzania (Jatropha B), respectively. The oils were produced from the seeds by the same mechanical oil extraction method. The chemicals used were analytical-grade reagents: potassium hydroxide and methanol, at 85% and 99.5% purity, respectively, and were purchased from local suppliers.

Biodiesel production process

The transesterification process was chosen for biodiesel production in this study, and the experiments were conducted using beakers as reactors. The reaction beaker was placed on a thermostatically controlled heating plate equipped with a magnetic stirrer whose maximum heating capacity is 420 °C and maximum agitation is 2000 rpm. A thermometer was fixed by the use of a retort stand and immersed in the beaker for verifying the temperature of the heating plate. The free fatty acid content was determined using the simple titration method as reported in the literature.[8]

The free fatty acid content of jatropha oils in this study was initially found to be 5.6% for Jatropha A and 5.8% for Jatropha B. An acid esterification process followed by the transesterification method was then used to prepare the *Jatropha curcas* methyl esters. The acid pre-treatment process was used to convert free fatty acids to esters using an acid catalyst (H_2SO_4) to reduce the free fatty acid concentration of the oil to below 1%. The acid catalyst (w/w of oil: 0.5% H_2SO_4) and a 6:1 methanol: oil molar ratio was used and the mixture stirred at 500 rpm for 1 h at 50 ºC. The mixture was allowed to settle, after which the pre-treated oil was collected and purified.[9] After acid esterification, transesterification was carried out at the following standard conditions: 6:1 methanol/oil molar ratio (mol/mol), 1.0 wt% potassium hydroxide, 55–60 °C reaction temperature, 400 rpm agitation speed and 60 min reaction time.[10] After the reaction and settling, the biodiesel (methyl esters) was collected and purified. The final product – the biodiesel – was a clear, light yellow liquid.

Measurement of biodiesel properties

The fuel properties of the samples were tested according to the standards shown in Table 1. All experiments were repeated at least twice and the results recorded when the standard deviation was less than 5%.

Table 1: Methods for determination of fuel properties

Property	Units	Test method
Cetane number	–	ASTM D613
Heating value	MJ/kg	ASTM D240
Density	kg/m³	ASTM D941
Viscosity	mm²/s	ASTM D445
Flash point	°C	ASTM D93-94
Water content	%	ISO 12937
Acid value	mg KOH/g	ASTM D974
Cloud point	°C	ASTM D2500
Lubricity	μm	ISO 12156
Oxidation stability	h	EN 14112

The fatty acid composition of biodiesel was analysed using a gas chromatograph mass spectrometer (Agilent 6890N) coupled to an inert mass-selective detector (Agilent 5973). The percentage composition of the individual components was obtained from electronic integration measurements using flame ionisation detection. All relative percentages determined by gas chromatography for each fatty acid sample are the means of three runs.

Emissions and combustion analysis

An Audi, 1.9-litre, turbocharged direct injection (TDI) compression ignition engine (Figure 1) was used to study engine performance. The engine is a four-stroke, four-cylinder, water-cooled diesel engine. The engine's technical specifications are given in Table 2.

Table 2: Details of the experimental engine used

Engine model	Audi, 1.9 L, turbocharged direct injection
Capacity	1896 cm³
Bore	79.5 mm
Stroke	95.5 mm
Compression ratio	19.5:1
Maximum power	66 kW, at 4000 rpm
Maximum torque	202 Nm, at 1900 rpm
Fuel system	Direct injection with electronic distributor pump

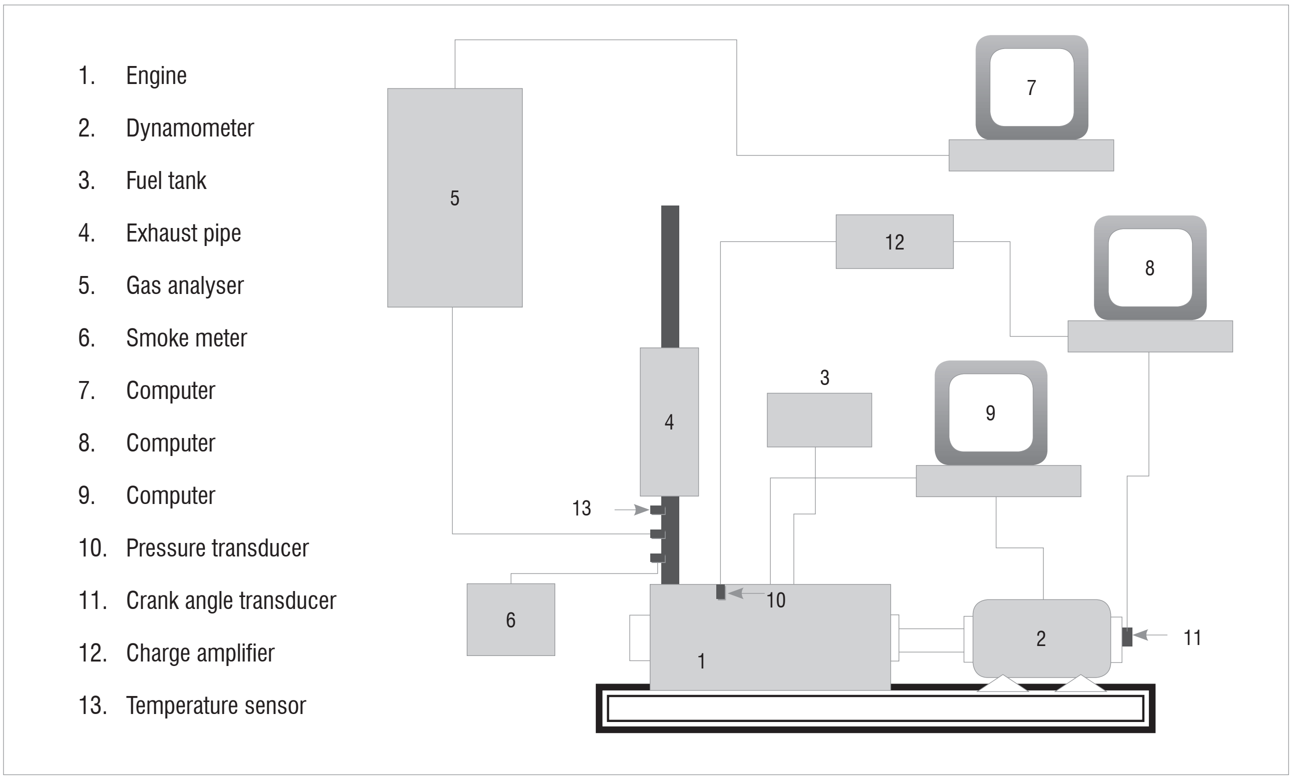


Figure 1: Engine test experimental set-up.

Figure 1 depicts the equipment connections for the engine test. A pressure transducer (10) was installed in one of the piston cylinders. Cylinder pressure signals from the pressure transducer were amplified by a charge amplifier (12) and connected to the SMETech COMBI-PC indication system (8) for data acquisition. The data acquisition system was externally triggered 1024 times in one revolution by an incremental crank angle transducer-optical encoder (11). Fuel was introduced from a fuel tank (3) equipped with a flow measurement system. During fuel switching, the fuel tank was drained from the engine fuel filter, new fuel was introduced into the tank until the fuel filter was full and the engine was then started and allowed to run for a few minutes to clear fuel lines and stabilise. Emission was measured by an Horiba emission analyser system (5) equipped with analyser modules NDIR (AIA-23), H.FID (FIA-22) and H.CLDC (CLA-53M) for measuring CO, CO_2 and unburnt hydrocarbons; THC; and NO_x, respectively, and by a smoke meter (6) connected before the oxidative converter at the engine exhaust pipe (4). The system was connected to the computer (7) and emission data was recorded.

Heat release calculations

During the engine test, the engine was run at a constant speed (3000 rpm) at different loads, from a low idle to 100% load at intervals of 25% of full load. Heat release was subsequently calculated from cylinder pressure and crank angle readings and emission data was then collected. The net heat release was calculated by a computer program using Equation 1 from pressure data (with respect to the crank angle) and cylinder geometry (with respect to the crank angle). The analysis was derived from the first law of thermodynamics for an open system which is quasistatic.[11]

$$\frac{dQn}{d\theta} = \frac{1}{\gamma-1}\left(\gamma p\frac{dV}{d\theta} + V\frac{dp}{d\theta}\right) \qquad \text{Equation 1}$$

where:

$\frac{dQn}{d\theta}$ = net heat-release rate (J/degree crank angle) and

γ = the ratio of specific heats, $\frac{C_p}{C_v}$

V and $\frac{dV}{d\theta}$ terms are calculated as shown in Equations 2 and 3, respectively:

$$V = V_c + A \times r\left[1 - \cos\left(\frac{\pi\theta}{180}\right) + \frac{1}{\lambda}\left\{1-\sqrt{1-\lambda^2\sin^2\left(\frac{\pi\theta}{180}\right)}\right\}\right]$$

Equation 2

and

$$\frac{dV}{d\theta} = \frac{\pi\theta}{180} \times r\left[\sin\frac{\pi\theta}{180} + \frac{\lambda^2\sin^2\left(\frac{\pi\theta}{180}\right)}{2 \times \sqrt{1-\lambda^2\sin^2\left(\frac{\pi\theta}{180}\right)}}\right] \qquad \text{Equation 3}$$

But $\lambda = \frac{1}{r}$ and $A = \frac{\pi}{4}D^2$

where *l* is the connecting rod length, *r* is the crank radius, *D* is the cylinder bore and V_c is the clearance volume.

Results and discussion

Biodiesel properties

The biodiesel from *Jatropha curcas* (Jatropha A and B) had a fatty acid composition as depicted in Table 3. As seen, the profiles for the two biodiesels are different. Table 4 summarises the properties of methyl esters obtained from the oils, which were found to be within the limits of biodiesel standards (ASTM D6751 and EN 14214), although they were different between the two biodiesels. The properties of fossil diesel are included for comparison purposes.

Table 3: Fatty acid composition of *Jatropha curcas* samples

Fatty acid composition (wt. %)	Jatropha A	Jatropha B
Palmitic methyl ester (C16/0)	15.3	16.1
Palmitoleic methyl ester (C16/1)	1.1	0.8
Stearic methyl ester (C18:0)	6.4	6.8
Oleic methyl ester (C18/1)	40.1	40.8
Linoleic methyl ester (C18/2)	36.9	34.6
Linolenic methyl ester (C18/3)	0.2	0.9
Saturated fatty acid methyl esters	21.7	22.9
Monounsaturated fatty acid methyl esters	41.2	41.6
Polyunsaturated fatty acid methyl esters	37.1	35.5

Table 4: Fuel properties of biodiesels Jatropha A and B, diesel and biodiesel standards

Property	Units	Jatropha fuel		Diesel	ASTM D6751	EN 14214
		A	B			
Density @ 15 °C	kg/m³	855	876	832	–	860-900
Viscosity @ 40 °C	mm²/s	4.33	4.37	3.01	1.9-6.0	3.5-5.0
Acid value	mgKOH/g	0.09	0.10	–	0.8 maximum	0.5 maximum
Flash point	°C	186	188	–	130 minimum	101 minimum
Heating value	MJ/kg	37.82	37.60	42.63	–	–
Lubricity	µm	218	222	–	–	–
Cetane number	–	56.54	55.13	54.60	47 minimum	–
Cloud point	°C	1	2	-16	–	–
Pour point	°C	-3	-2	-19	–	–
Water content	ppm	460	498	50	500 maximum	500 maximum
Sulphur content	ppm	0.0025	0.0020	45		
Oxidation stability at 110 °C	h	9.65	6.51	–	–	minimum

Engine emissions

Engine emissions obtained experimentally were plotted against load as described below.

Oxides of nitrogen

Figure 2 depicts the oxides of nitrogen (NO_x) emissions of the fuel samples tested at different load conditions. In general, the formation of NO_x was affected by the peak flame temperature, the high burning gas temperature, the ignition delay and the content of nitrogen and oxygen available in the reaction mixture.[12] It was observed that as the load increases, the in-cylinder temperature also increases and thus higher absolute NO_x (ppm) formation results. Higher cylinder pressure could contribute to increased NO_x emissions, as a result of the increased peak combustion temperature at higher engine loads. These results correspond to those in the literature.[13,14] Fossil diesel (D2) exhibited slightly higher NO_x emissions compared with Jatropha A, especially at lower engine loads; however, Jatropha B had the highest overall emission of NO_x. A possible reason for this finding may be the slow burning of the more viscous biodiesel, because it spent longer in the high temperature zone.[15] Moreover, the NO_x emission increase can be associated with the oxygen content in the fuel samples as fuel with oxygen may provide additional oxygen for the formation of NO_x.[16]

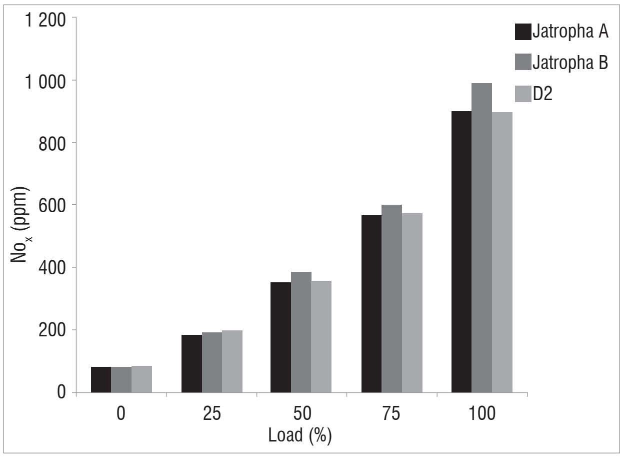


Figure 2: NO_x emissions for the different fuel samples at different loads.

Total hydrocarbons

The variation of THC emissions with load is shown in Figure 3. It can be observed that THC emissions are a maximum on idling, whereas lower THC values are obtained at 50% and 75% loads at steady engine speed. A similar trend was also reported in the literature, that is, high THC emissions were observed at lower loads and maximum loads.[17] The small difference between D2 and the jatropha fuels in THC emissions could possibly be attributed to the higher fuel supply for a given load for jatropha, which produces slower combustion times and counteracts the possible benefit of the presence of fuel-borne oxygen in enhancing the combustion process of the jatropha fuels.[18,19]

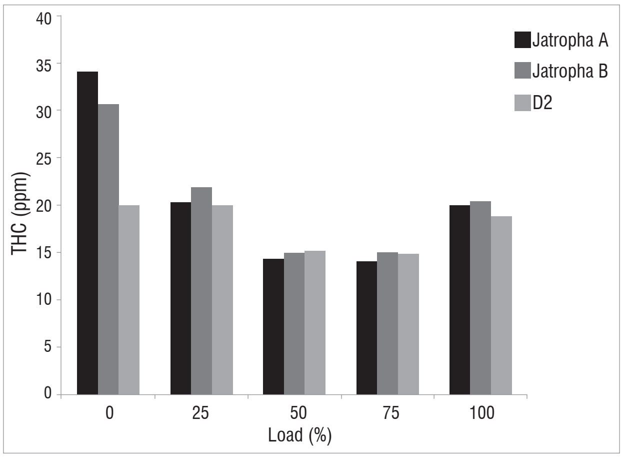


Figure 3: Total hydrocarbon (THC) emissions for the different fuel samples at different loads.

Carbon monoxide

The formation of CO with increasing load is illustrated in Figure 4. The figure shows that CO emission decreases with increasing engine load. High CO emissions were observed at lower loads, with the lowest emissions recorded at 75% of maximum engine load. CO is a product of incomplete combustion; thus at higher engine loads, the high combustion temperature promotes more complete combustion and hence fewer CO emissions. These results are comparable with those reported in the literature.[14,19] The relatively poor atomisation and lower volatility of fuel samples is responsible for this trend. In addition, at lower loads, CO emissions are higher because of incomplete combustion, whereas at 100% load, CO emissions are slightly increased compared with at 75% load, as a result of the local presence of a richer mixture in the combustion chamber.[20]

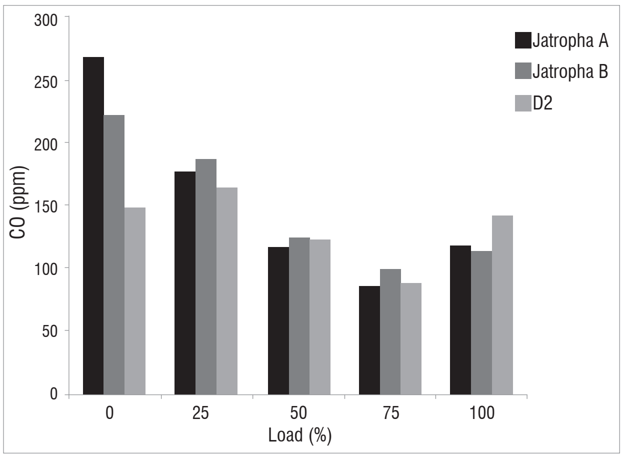


Figure 4: Carbon monoxide (CO) emissions for the different fuel samples at different loads.

Carbon dioxide

Fuel derived from vegetable oils results in fewer CO_2 emissions when used to run diesel engines because plants absorb CO_2 in a sustainable fashion during growth.[21,22] Figure 5 illustrates the CO_2 emissions at different loads for the two biodiesels and fossil diesel. It can be observed that CO_2 emissions increase as the load increases at the same engine speed. Oxygen present in biodiesel supports combustion which leads to higher CO_2 emissions at low loads as a result of complete combustion. The finding that diesel exhibits lower CO_2 emissions has been reported previously.[23] At higher loads, however, poor fuel atomisation and a lower combustion duration override the advantages of bound oxygen in Jatropha A and B, thus incomplete combustion results.

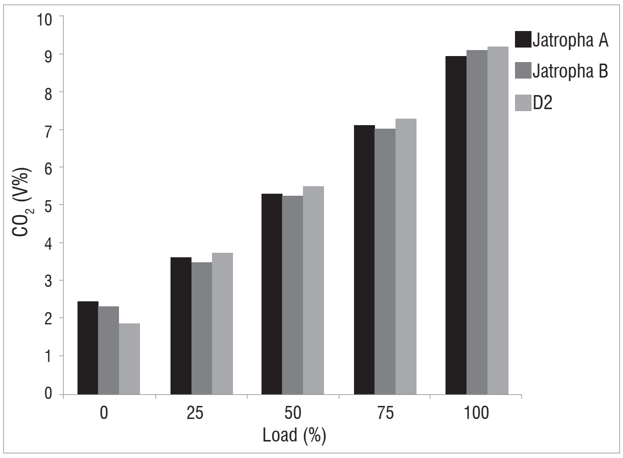


Figure 5: Carbon dioxide (CO_2) emissions for the different fuel samples at different loads.

Smoke

Smoke emissions with increasing load are shown in Figure 6. Diesel fuel exhibited the highest smoke emissions at all loads, whilst Jatropha B exhibited the lowest emissions under the same load conditions. Maximum smoke emission values were observed at 100% load for all fuels. The presence of oxygen in Jatropha A and B samples could explain their lower emission values.[13,24]

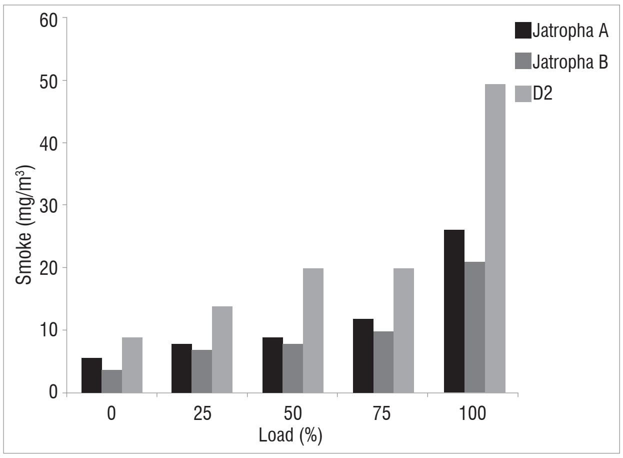


Figure 6: Smoke emissions for the different fuel samples at different loads.

Combustion analysis

Variation in injection timing exerts a major influence on engine performance and exhaust emissions. To understand and optimise the combustion process, a careful analysis of cylinder pressure and heat release was performed, which furnished precise information about the combustion of biodiesel. A TDI compression ignition engine was used to undertake an analysis of the combustion characteristics of the fuel samples tested.

Cylinder pressure

Figure 7 indicates the peak cylinder pressure for different fuels at different engine load conditions. Figure 8 illustrates the variations of cylinder pressure with crank angle degree for the fuels at full load and a constant engine speed of 3000 rpm. Similar trends were obtained at other loads and differed only in the magnitude of the pressure and the corresponding crank angle at which the pressure appeared. In a compression ignition engine, the cylinder pressure depends on the burned fuel fraction during the pre-mixed burning phase, that is, the initial stage of combustion. Cylinder pressure characterises the ability of the fuel to mix well with air and burn. A high peak pressure and maximum rate of pressure rise correspond to a large volume of fuel burned in the pre-mixed combustion stage.[25]

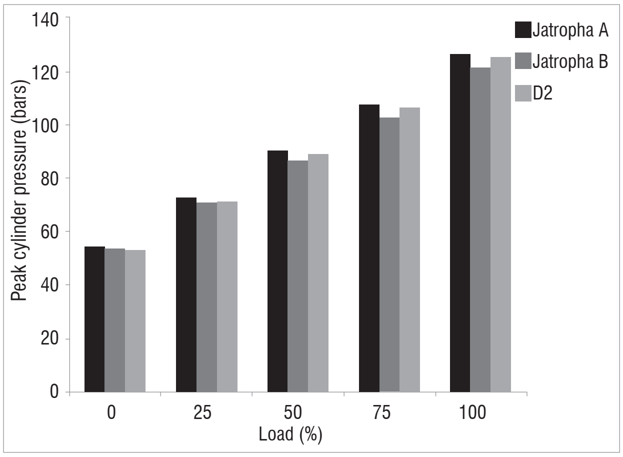


Figure 7: Peak cylinder pressure for the different fuel samples at different loads.

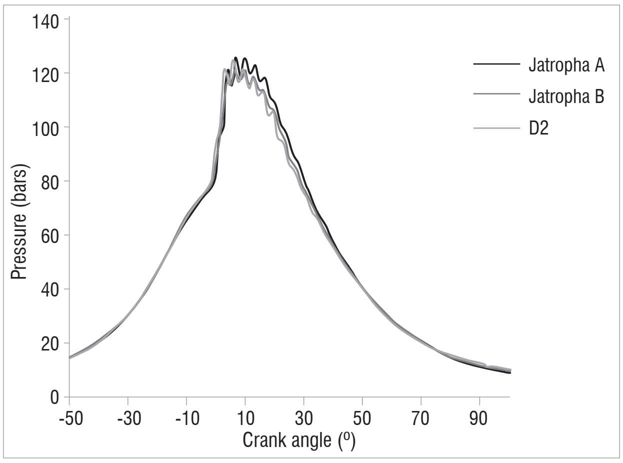


Figure 8: Cylinder pressure variation with crank angle at full load.

It can be noted that Jatropha B produced the lowest peak cylinder pressure while Jatropha A produced the highest at all loads. All fuels exhibited peak cylinder pressures between crank angles of -10° and 10°, with lower load levels exhibiting peak cylinder pressure at negative crank angles and vice versa. It was also observed that the cylinder pressure increases as the load increases. This increase in pressure is to be expected in a TDI diesel engine and is caused by the increased amount of injected fuel and charge pressure in the air intake manifold at higher loads.

The temperature and volume flow of the exhaust gases that pass through the turbo charger influence the characteristics of air flow in the intake manifold, because of the increased spinning speed of the turbocharger turbine and the heat exchange between exhaust gases and the charged air. The amount of air forced into the intake manifold increases as the load increases, because the high temperature, high velocity exhaust gas that passes through the turbocharger turbine causes a rise in the velocity of intake air, which forces more air into the cylinders. The increased amount of air causes larger amounts of fuel to be injected into the cylinder, which in turn increases the cylinder pressure.

Combustion of Jatropha A resulted in the highest peak pressures probably because of the increase in ignition delay associated with using Jatropha A. This ignition delay increases the amount of fuel burned within the pre-mixed burning phase, causing high peak pressures and a high rate of pressure rise. Jatropha A's lower viscosity (compared with Jatropha B) and the contribution of intrinsic oxygen (compared with D2) may also be the reason for the high cylinder pressure. The high viscosity of Jatropha B is the reason for its lower peak cylinder pressure, as fuel spraying is affected by viscosity. Complete combustion depends to a large extent on fuel–air mixing. More complete combustion results in a higher cylinder pressure than incomplete combustion.[13,14] Diesel exhibited a slightly lower peak cylinder pressure than Jatropha A; this finding could be a result of oxygen–fuel mixing, which is more efficient in fuel that contains intrinsic oxygen.[16,25]

Heat release

Figure 9 depicts peak heat release rates for the different fuels. Figure 10 illustrates the variations of heat release with crank angle degree for the fuels at full load conditions and a constant engine speed of 3000 rpm. Similar trends were observed at other loads and differed only in the magnitude of heat release and the corresponding crank angle at which the heat release occurred. For all fuels, the peak heat release increased as the load increased, possibly as a result of high temperature and high cylinder pressure, better fuel–air mixing, and higher flame velocity at higher loads. All fuels experienced rapid pre-mixed burning followed by diffusion combustion, as is typical for naturally aspirated engines. After the ignition delay period, the pre-mixed fuel–air mixture burns rapidly, releasing heat at a very rapid rate, after which diffusion combustion takes place, during which the burning rate is controlled by the availability of the combustible fuel–air mixture. It can be seen that Jatropha A and B showed an improvement in the heat release rate during the pre-mixed combustion period. The presence of oxygen in these fuels decreased their cetane number and increased the ignition delay period. Therefore, while the engine was running with biodiesel, increased accumulation of fuel during the relatively longer delay period resulted in a higher rate of heat release. Even though diesel exhibited a higher heating value and lower viscosity, the intrinsic oxygen property of the other fuels influenced the heat release results obtained. Because of the shorter delay period for diesel, its maximum heat release rate occurred earlier in comparison with the jatropha fuels. Jatropha A's higher heat release compared with Jatropha B is probably because of its higher heating value and lower viscosity.

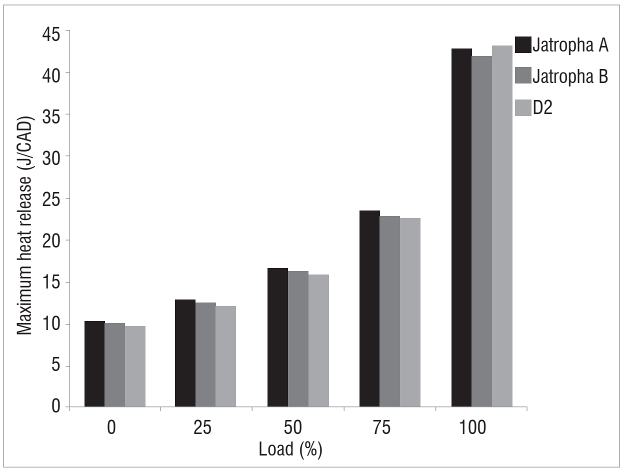


Figure 9: Maximum rate of heat release at different engine loads.

Effect of geographical location on jatropha oil properties

A good explanation of why different regions produce jatropha oils with different properties was given by Zhang et al.[26] While studying the genetics of the plant, they discovered that the plant tends to adapt to its specific habitat at a specific altitude and latitude, and changes its characteristics in the process. They further explained that these geographical environments have different ecological factors that result in the exposure of the plants to different selection pressures, eventually leading to genetic differentiation among plants from different regions. The same observations were made by Kaushik et al.[27] in their genetic study of the seed traits and oil content of *Jatropha* from different locations in India. Kaushik et al. explained that the variations in jatropha seeds with respect to their morphological characteristics could be because the species grows over a wide range of rainfall conditions, temperatures and soil types. Ovando-Medina et al.[28] also concluded that *Jatropha* trees differ according to their geographical origin. Their study, which was conducted on trees from different regions in Mexico, indicated a correlation between chemical properties and geographical location. Similar observations were made by Popluechai et al.[29] who reported that jatropha oil from different parts of the world (Africa, Asia and South America) exhibited different genetic and physicochemical characteristics. Thus, different regions with different geographical and environmental conditions produce trees with different characteristics.

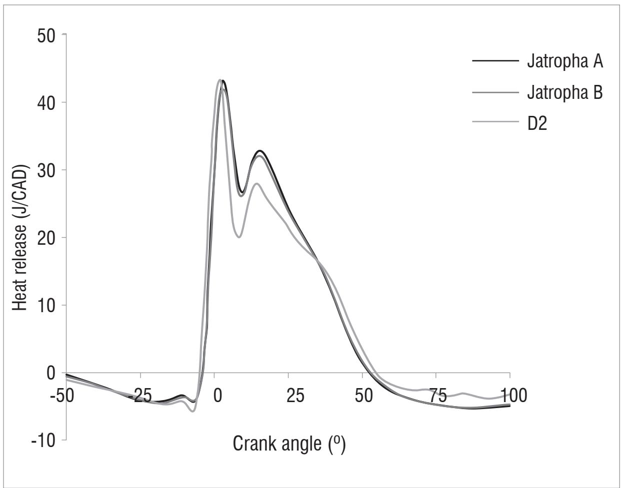


Figure 10: Heat release variation with crank angle at full load.

Conclusion and recommendations

A two-step acid–alkaline catalytic transesterification process was used to prepare two samples of jatropha oil from different countries in East Africa. The samples produced were compared with each other and to fossil diesel in terms of engine emissions and combustion characteristics. It was observed that the biodiesels had different emission and combustion characteristics, although both performed better than the fossil fuel. It was not clear which biodiesel performed best overall because Jatropha A was better in some cases and Jatropha B in others. The difference between the biodiesels is related to the respective geographical locations of the original oils. Different geographical locations produce plants with varying characteristics as a result of the different ecological and environmental conditions to which the plants genetically adapt.

Acknowledgements

The author acknowledges the financial and logistical assistance provided by Tshwane University of Technology (South Africa) and Moi University (Kenya).

References

1. Makkar HPS, Becker K. *Jatropha curcas*, a promising crop for the generation of biodiesel and value-added coproducts. Eur J Lipid Sci Tech. 2009;111:15. http://dx.doi.org/10.1002/ejlt.200800244
2. Treese JV, Paper M, Li Y, Wang Q, Moore K. *Jatropha curcas* cultivation and use as a biodiesel feedstock. Florida: University of Florida; 2010.
3. Pramanik K. Properties and use of *Jatropha curcas* oil and diesel fuel blends in compression ignition engine. Renew Energ. 2003;28(2):239–248. http://dx.doi.org/10.1016/S0960-1481(02)00027-7
4. Emil A, Yaakob Z, Satheesh Kumar M, Jahim J, Salimon J. Comparative evaluation of physicochemical properties of jatropha seed oil from Malaysia, Indonesia and Thailand. J Am Oil Chem Soc. 2010;87(6):689–695. http://dx.doi.org/10.1007/s11746-009-1537-6
5. Lu H, Liu Y, Zhou H, Yang Y, Chen M, Liang B. Production of biodiesel from *Jatropha curcas* L. oil. Comput Chem Eng. 2009;33(5):1091–1096. http://dx.doi.org/10.1016/j.compchemeng.2008.09.012
6. Foidl N, Foidl G, Sanchez M, Mittelbach M, Hackel S. *Jatropha curcas* L. as a source for the production of biofuel in Nicaragua. Bioresource Technol. 1996;58(1):77–82. http://dx.doi.org/10.1016/S0960-8524(96)00111-3
7. Lalas S, Tsaknis J. Characterization of *Moringa oleifera* seed oil variety "Periyakulam 1". J Food Compos Anal. 2002;15(1):65–77. http://dx.doi.org/10.1006/jfca.2001.1042
8. Knothe G, Van Gerpen JH, Krahl J. The biodiesel handbook. Champaign, IL: AOCS Press; 2005.
9. Ramadhas AS, Jayaraj S, Muraleedharan C. Biodiesel production from high FFA rubber seed oil. Fuel. 2005;84(4):335–340. http://dx.doi.org/10.1016/j.fuel.2004.09.016
10. Rashid U, Anwar F, Moser BR, Knothe G. *Moringa oleifera* oil: A possible source of biodiesel. Bioresource Technol. 2008;99(17):8175–8179. http://dx.doi.org/10.1016/j.biortech.2008.03.066
11. Ren Y, Huang Z, Miao H, Di Y, Jiang D, Zeng K, et al. Combustion and emissions of a DI diesel engine fuelled with diesel-oxygenate blends. Fuel. 2008;87(12):2691–2697. http://dx.doi.org/10.1016/j.fuel.2008.02.017
12. Heywood JB. Internal combustion engine fundamentals. New York: McGraw-Hill; 1998.
13. Purushothaman K, Nagarajan G. Performance, emission and combustion characteristics of a compression ignition engine operating on neat orange oil. Renew Energ. 2009;34(1):242–245. http://dx.doi.org/10.1016/j.renene.2008.03.012
14. Purushothaman K, Nagarajan G. Experimental investigation on a C.I. engine using orange oil and orange oil with DEE. Fuel. 2009;88(9):1732–1740. http://dx.doi.org/10.1016/j.fuel.2009.03.032
15. Agarwal D, Sinha S, Agarwal AK. Experimental investigation of control of NOx emissions in biodiesel-fueled compression ignition engine. Renew Energ. 2006;31(14):2356–2369. http://dx.doi.org/10.1016/j.renene.2005.12.003
16. Ramadhas AS, Muraleedharan C, Jayaraj S. Performance and emission evaluation of a diesel engine fueled with methyl esters of rubber seed oil. Renew Energ. 2005;30(12):1789–1800. http://dx.doi.org/10.1016/j.renene.2005.01.009
17. Shi X, Pang X, Mu Y, He H, Shuai S, Wang J, et al. Emission reduction potential of using ethanol-biodiesel-diesel fuel blend on a heavy-duty diesel engine. Atmos Environ. 2006;40(14):2567–2574. http://dx.doi.org/10.1016/j.atmosenv.2005.12.026
18. Aliyu B, Shitanda D, Walker S, Agnew B, Masheiti S, Atan R. Performance and exhaust emissions of a diesel engine fuelled with *Croton megalocarpus* (musine) methyl ester. Appl Therm Eng. 2011;31(1):36–41. http://dx.doi.org/10.1016/j.applthermaleng.2010.07.034
19. Banapurmath NR, Tewari PG, Hosmath RS. Performance and emission characteristics of a DI compression ignition engine operated on honge, jatropha and sesame oil methyl esters. Renew Energ. 2008;33(9):1982–1988. http://dx.doi.org/10.1016/j.renene.2007.11.012
20. Cheng CH, Cheung CS, Chan TL, Lee SC, Yao CD, Tsang KS. Comparison of emissions of a direct injection diesel engine operating on biodiesel with emulsified and fumigated methanol. Fuel. 2008;87(10–11):1870–1879. http://dx.doi.org/10.1016/j.fuel.2008.01.002
21. Demirbas A. Biodiesel: A realistic fuel alternative for diesel engines: London: Springer Verlag; 2008.
22. Balat M, Balat H. A critical review of bio-diesel as a vehicular fuel. Energ Convers Manage. 2008;49:2727–2741. http://dx.doi.org/10.1016/j.enconman.2008.03.016
23. Mbarawa M. Performance, emission and economic assessment of clove stem oil-diesel blended fuels as alternative fuels for diesel engines. Renew Energ. 2008;33(5):871–882. http://dx.doi.org/10.1016/j.renene.2007.06.009

24. Chen H, Wang J, Shuai S, Chen W. Study of oxygenated biomass fuel blends on a diesel engine. Fuel. 2008;87(15–16):3462–3468. http://dx.doi.org/10.1016/j.fuel.2008.04.034

25. Agarwal AK. Biofuels (alcohols and biodiesel) applications as fuels for internal combustion engines. Prog Energ Combust Sci. 2007;33(3):233–271. http://dx.doi.org/10.1016/j.pecs.2006.08.003

26. Zhang Z, Guo X, Liu B, Tang L, Chen F. Genetic diversity and genetic relationship of *Jatropha curcas* between China and Southeast Asian revealed by amplified fragment length polymorphisms. Afr J Biotechnol. 2011;10(15):2825–2832.

27. Kaushik N, Kumar K, Kumar S, Kaushik N, Roy S. Genetic variability and divergence studies in seed traits and oil content of jatropha (*Jatropha curcas* L.) accessions. Biomass Bioenerg. 2007;31(7):497–502. http://dx.doi.org/10.1016/j.biombioe.2007.01.021

28. Ovando-Medina I, Espinosa-García F, Núñez-Farfán J, Salvador-Figueroa M. Genetic variation in Mexican *Jatropha curcas* L. Estimated with seed oil fatty acids. J Oleo Sci. 2010.

29. Popluechai S, Breviario D, Mulpuri S, Makkar HPS, Raorane M, Reddy AR, et al. Narrow genetic and apparent phenetic diversity in *Jatropha curcas*: Initial success with generating low phorbol ester interspecific hybrids. Nature Precedings. 2009:2782.

High biomass yielding winter cover crops can improve phosphorus availability in soil

AUTHORS:
Ernest Dube[1]
Cornelius Chiduza[2]
Pardon Muchaonyerwa[3]

AFFILIATIONS:
[1]Agricultural Research Council – Small Grain Institute Production Systems, Bethlehem, South Africa
[2]Department of Agronomy, University of Fort Hare, Alice, South Africa
[3]Soil Science, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

CORRESPONDENCE TO:
Ernest Dube

EMAIL:
dubee@arc.agric.za

POSTAL ADDRESS:
Agricultural Research Council – Small Grain Institute Production Systems, Lindley Road, Bethlehem 9700, South Africa

DATES:
Received: 06 May 2013
Revised: 26 Oct. 2013
Accepted: 29 Oct. 2013

KEYWORDS:
conservation agriculture; cover crop biomass; no-till; phosphorus fertiliser; soil organic matter

HOW TO CITE:
Dube E, Chiduza C, Muchaonyerwa P. High biomass yielding winter cover crops can improve phosphorus availability in soil. S Afr J Sci. 2014;110(3/4), Art. #2013-0135, 4 pages. http://dx.doi.org/10.1590/sajs.2014/20130135

We investigated the effects of high biomass yielding winter cover crops, namely grazing vetch (*Vicia dasycarpa* L.) and oats (*Avena sativa* L.), on soil phosphorus (P) availability in low fertiliser input maize-based conservation agriculture systems. Soil samples were collected from the 0–50-mm depth of experimental plots after 4 years of maize–winter cover crop rotations. A sequential fractionation scheme was used to separate total soil P into labile, moderately labile and non-labile organic P (P_o) and inorganic (P_i) pools. Labile P pools included microbial biomass-P as well as P_i and P_o pools extracted using 0.5 M $NaHCO_3$ and 1.0 M HCl. The non-labile P pools were humic-P and 1.0 M H_2SO_4 extracted P. Soil on the maize–winter cover crop rotations had higher HCl-P_i and total P than the soil on the maize–fallow rotation. The cover crops had no significant ($p>0.05$) effect on $NaHCO_3$-P_o, $NaHCO_3$-P_i, HCl-P_o, fulvic acid-P and recalcitrant H_2SO_4-P fractions. Non-application of fertiliser increased accumulation of humic-P on the maize–oats rotation. Cover crop biomass input explained 73% of the variations in microbial biomass-P and 33% of variations in total labile P. Phosphorus concentration of young maize plants was significantly increased by the cover crops, with a positive correlation to HCl-P_i ($r_s=0.90$). This contribution from winter cover crops to P availability in the surface soil suggests that, in the long term, fertiliser P could be reduced in such systems.

Introduction

Phosphorus (P) is an essential macronutrient required in almost every aspect of plant functions. It is a vital component of compounds required to build proteins, plant structure, seed yield and genetic transfer. Symptoms of P deficiency in maize include stunted growth, delayed maturity, purplish hues, poor root development and reduced yield potential. Phosphorus is the second most important fertiliser applied to soils for improving soil fertility after nitrogen (N) in maize production. Cultivated soils in South Africa are deficient in P and fertiliser compounds have to be supplied to soils in large quantity to meet maize P requirements. Meanwhile, it is reported that the world's P stocks are dwindling and P production will not be able to meet half the world's needs by the year 2050.[1] Global P fertiliser prices are on the increase[1,2] and eating into the profits of farmers. It is becoming increasingly imperative for maize farmers to adopt low-cost farming strategies that increase P availability in the soil, conserve soil P and optimise its use.

Conservation agriculture (CA) is an important farming practice for conserving soil resources, including P, through retention of crop residues and minimising soil erosion.[3] In CA, the practices of crop rotation, no till and permanent cover through crop residues and cover crops contribute immensely to soil quality and nutrient dynamics. A number of winter cover crops, which are legumes or small grains, can be grown between regular grain crop production periods for the purpose of protecting and improving the soil through their high biomass production.[4] Through their extensive root systems, leguminous winter cover crops can explore subsoil nutrient pools, whereas grass species can increase P uptake by both the cover crop and the succeeding crop through enhancement of viable mycelia of mycorrhizal fungi in soils.[5,6] Almost no P is added to the soil by winter cover crops; they only take up P from the soil solution and return it to the same soil. However, high biomass yielding winter cover crops can increase surface organic matter significantly.[7]

In warm temperate regions such as the Eastern Cape Province of South Africa, planting a winter cover crop before the summer maize crop is possible as an entry point into CA, provided irrigation water is available.[7,8] Grazing vetch (*Vicia dasycarpa*) and oats (*Avena sativa*) are examples of fast-growing, winter hardy cover crops which can provide high biomass (> 6 t/ha) in this system.[8] The maize is planted immediately after cover crop termination onto winter cover crop residue mulches. Decomposition of the winter cover crop residues and subsequent mineralisation of organic P (P_o) plays an essential role in P-cycling and maintenance of plant-available P in this system.[9] The major source of P in unfertilised low P soils is P_o.[9,10] When no winter cover crops are used, biomass production tends to be lower, reducing the size of the P_o pool. However, where high amounts of organic matter are generated in situ from winter cover crops, P may be temporarily immobilised in the organic matter accumulated on the soil surface, especially if the C:P ratio is greater than 300:1.[11] There is therefore a need to investigate the effects of winter cover crops on maize P nutrition, especially during the early crop growth stages. Adequate P nutrition at the seedling stage in maize is critical because deficiency at this stage cannot be remedied by side-dressing as a result of a lack of P mobility in soils.

The size of the P_o pool that undergoes rapid mineralisation contributing to plant available P over at least one growing season, known as labile P, may be dependent on crop residue quality, soil and environmental characteristics, and the duration and type of the cropping system.[12,13] Bicarbonate extractable P ($NaHCO_3$-P) and microbial biomass-P fractions constitute labile P in the soil.[10,14] The soil microbial biomass may be considered as a reservoir of potentially plant-available nutrients, including P.[14] It is important to understand the effects of winter cover crops on these P pools in low fertiliser input CA systems for the development of effective fertiliser management strategies that maximise maize yield and profit. In this paper, we report the effects of winter cover crops on soil P pools and maize P nutrition during early growth in a low fertiliser input CA system.

Materials and methods

A long-term field experiment was established in 2007 on a research farm to study the effects of winter cover crops and fertiliser on biomass input, soil organic matter and maize yield under no-till and irrigation conditions.[8] The farm is located at 32°46′ S 26°50′ E and at 535 m above sea level in the Eastern Cape Province of South Africa. The soil type and climate of this research site has been described previously.[8] The field trial was a split plot design and the winter cover crops oats (cv. Sederberg) and grazing vetch (cv. Max) were planted in the main plots. Control plots with no winter cover crops (fallow) were also included. Subsequent to winter cover crop termination, plots were split and summer maize was planted at two fertiliser levels (with and without fertiliser). Fertiliser was never applied to the cover crops or the fallow plots. There were thus two factors in this experiment – type of cover crop mulch and fertiliser application – resulting in a 3×2 split plot design which was replicated three times. The six treatment combinations and the amount of P fertiliser applied are presented in Table 1.

Table 1: The treatments used in the study

Type of cover crop (main plots)	Maize (subplots)	Treatment
Oats	60 kg N and 30 kg P per ha applied to the maize	1
	No fertiliser applied	2
Vetch	60 kg N and 30 kg P per ha applied to the maize	3
	No fertiliser applied	4
Fallow	60 kg N and 30 kg P per ha applied to the maize	5
	No fertiliser applied	6

Detailed agronomic management of the field trial has been described previously.[8] Twelve random soil samples were collected from the organic matter rich depths (0–50 mm) of the plots at the beginning of the fourth year using a precision auger (70 mm diameter) after cover crop termination and before maize planting. Soil samples were collected from the inner two thirds of each plot and the soil from each plot was bulked to form one sample. The samples were air dried and sieved (<2 mm) to remove coarse fragments and roots. At 6 weeks after maize planting, six maize plants per plot (two from each of rows 2, 7 and 8) were sampled by cutting near the soil surface. The plants were oven dried to a constant weight at 65 °C and dry matter was determined before the sample was ground (<2 mm).

Soil P was separated into labile, moderately labile and non-labile organic and inorganic pools following the sequential fractionation scheme.[12,15] In this method, the labile P pool was extracted using 0.5 M $NaHCO_3$ at pH 8.5, while the microbial biomass-P was determined through a chloroform ($CHCl_3$) fumigation–extraction technique.[8] The moderately labile pool was extracted with 1.0 M HCl, followed by 0.5 M NaOH. The NaOH extract was acidified with concentrated HCl to separate the non-labile humic acid P fraction from the moderately labile fulvic acid P fraction. Residue from the NaOH was ashed at 550 °C for 1 h and dissolved in 1.0 M H_2SO_4 to determine the highly resistant, non-labile P fraction. Total P in all extracts was measured after persulphate digestion.[16] Organic P in the extracts was calculated as the difference between total P and inorganic P (P_i). Plant tissue P concentration of maize plants was extracted using a wet digestion procedure with H_2SO_4 and H_2O_2.[17] Phosphorus concentrations in all extracts were analysed by continuous flow analysis using the molybdenum blue colourimetric method on a Skalar San Plus System (Breda, the Netherlands).[18] All data were subjected to an analysis of variance as a split plot design to test the effects of the winter cover crops and the fertiliser using GenStat Release 12.1 statistics software. Separation of means was done by using the least significant difference at a 5% level of significance. Biomass inputs from the winter cover crops in this field trial have been reported previously.[8] A Spearman's rank correlation was used in the current study to analyse the biomass data.

Results and discussion

Winter cover crop type × fertiliser interaction effect on total soil P was not significant ($p>0.05$). Cover crop type effects were significant ($p<0.05$), that is, total P in soils was higher on maize–oats (619±32 mg/kg) and maize–grazing vetch (634±41 mg/kg) rotations than in soils on maize–fallow rotation (524±27 mg/kg). This finding implies that winter cover crop residues can be a source of P in the low P fertiliser input no-till systems. This P is mined from larger soil volumes through their extensive root systems. When the cover crops are terminated, the crop residue P contributes to the total P levels in the surface soils of no-till systems. This soil P, however, occurs in different pools which vary in the level of P availability to plants. Fertiliser effects on total soil P were not significant ($p>0.05$).

Winter cover crop type × fertiliser interaction effects on all the labile and moderately labile soil P pools were not significant ($p>0.05$) (Table 2). However, winter cover crop type had a significant ($p<0.05$) effect on microbial-P and HCl-P_i, but not on $NaHCO_3$-P_o, $NaHCO_3$-P_i, HCl-P_o and fulvic acid-P (Table 2). The soil on maize–oats rotation had higher microbial-P than the soil on either the maize–grazing vetch or maize–fallow rotation (Figure 1). Oats residues tend to have a slower decomposition rate than vetch residues.[7] In this case, most mineral soil P in the soil under oats residues may be converted into microbial biomass.

Bicarbonate is thought to solubilise P that is adsorbed on surfaces of crystalline P-compounds, carbonates and oxides of Fe and Al.[19] The lack of significant differences in bicarbonate-extractable P fractions ($NaHCO_3$-P_o, $NaHCO_3$-P_i) among the cover crop treatments could be caused by low contents of crystalline minerals like carbonates and oxides of Fe and Al[20] in the soil used for the study, which was a Haplic cambisol. $NaHCO_3$-P also includes soluble P, which is often detected using ion-exchange resins.[19]

Table 2: Summary of analysis of variance of the effects of winter cover crop type and fertiliser on phosphorus pools after 4 years of maize–winter cover crop rotation

	Labile to moderately labile organic (P_o) and inorganic (P_i) pools						Non-labile P_o pools	
	$NaHCO_3$-P_o	$NaHCO_3$-P_i	Microbial-P	HCl-P_o	HCl-P_i	Fulvic acid-P	Humic acid-P	H_2SO_4-P_o
Winter cover crop (main plot)	ns	ns	**	ns	**	ns	**	ns
Fertiliser (subplot)	*	ns	ns	ns	ns	ns	ns	ns
Winter cover crop × fertiliser	ns	ns	ns	ns	ns	ns	**	ns
Coefficient of variation (%)	16.3	16.2	19.3	16.9	18.4	17.3	16.9	5.3

*ns, not significant; *p<0.05; **p<0.01*

Soil on the maize–oats and maize–grazing vetch rotations had higher HCl-P_i than that on the maize–fallow rotation (Figure 1). This high level of the inorganic pool of HCl-P fraction in the cover crop treatments suggests that when the large biomass decomposes a greater proportion of the mineral P produced forms calcium phosphates,[20] and a portion is taken up by soil microbial biomass. Dissolution of these calcium phosphates makes moderately labile P (HCl-P) available to crops.[21]

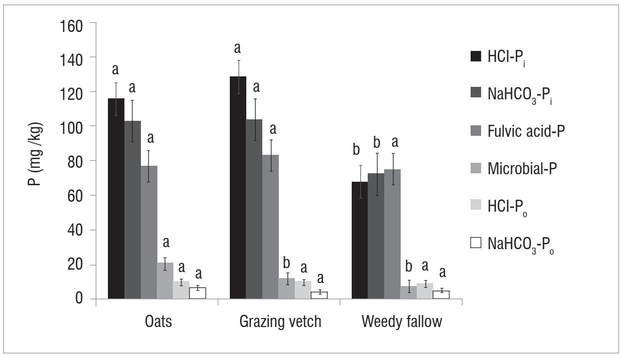


Figure 1: Effects of winter cover crop type on labile and moderately labile soil P pools.

Fertiliser had no significant ($p > 0.05$) effect on all the soil P pools except for $NaHCO_3$-P_o (Table 2) for which the fertilised maize rotations had a higher amount (6.78 mg/kg) than the non-fertilised ones (4.33 mg/kg). Winter cover crop × fertiliser interaction had a significant ($p < 0.05$) effect on humic-P, but not on H_2SO_4-P_o. Soil in the non-fertilised maize–oats rotation had the highest amount of humic acid-P while the fertilised maize–fallow had the lowest amount (Figure 2). Whereas humic-P in the maize–vetch rotation was not increased by fertiliser application, humic-P in the maize–oats and maize–fallow rotations was higher when unfertilised. Although humic-P and H_2SO_4-P are considered non-labile, their effects as a slow supply of plant available P over the long term could be significant. The high humic-P in the non-fertilised maize–oats rotation suggests an accumulation of humic acid fraction of organic matter which could be effectively locking up P over the short term. Decomposition of the humic fraction of soil organic matter proceeds at a slower rate than the material from which it was formed. A possible explanation for high humic-P under non-fertilised maize–oats rotation is the fact that N fertiliser application accelerates decomposition of the humic acid fraction to form HCl-P_i and/or microbial biomass-P. Application of fertiliser generally increases the soil organic matter mineralisation rate and reduces humic acid content.[22]

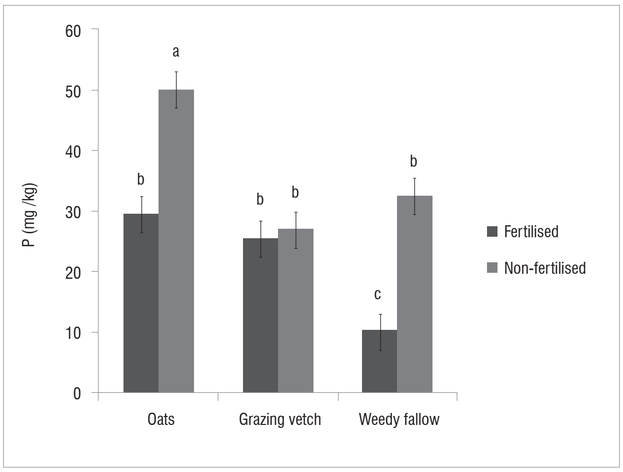


Figure 2: Interaction effects of winter cover crop type and fertiliser on humic acid-P.

Winter cover crop × fertiliser interaction effect on maize plant P concentration was not significant ($p > 0.05$). However, winter cover crop type effects were significant and the maize planted after fallow had a lower tissue P concentration (1138±72 mg/kg) than that planted after either oats (1650±88 mg/kg) or grazing vetch (1759±103 mg/kg). It is possible that cover crop residues release organic acids that compete for P sorption sites with orthophosphate, thereby resulting in increased amounts of P in the soil solution for crop uptake. Soluble plant litter P can also be released to the soil from cover crop residues as an initial flush of P_i with rainfall or irrigation for utilisation by the maize crop during early stages of growth.[23] Decomposition of the cover crop residues can further increase P availability by releasing CO_2, which forms H_2CO_3 in the soil solution, resulting in the dissolution of primary P-containing minerals.[24] Fertiliser effects on maize plant tissue P concentration were not significant ($p > 0.05$).

A Spearman's rank correlation analysis of the P data showed a strong positive relationship between HCl-P_i and the maize P concentration ($r_s = 0.90$) (Table 3). This result suggests that maize was dependent on these labile P_i pools for P nutrition during early growth. Correlations of maize tissue P concentration with microbial-P, HCl-P_o, $NaHCO_3$-P_o, $NaHCO_3$-P_i and fulvic acid-P were not significant (Table 3). Spearman's rank correlation coefficients between winter cover crop biomass input (summed over 4 years) and soil P pools are presented in Table 3. Total biomass accumulation (maize stover + winter cover crop biomass) significantly correlated with HCl-P_i, microbial P, total labile P and maize plant tissue P (Table 3). Winter cover crop biomass input was correlated with humic-P, whereas total biomass input was correlated with HCl-P_i (Table 3). This finding suggests that at least some component of the humic-P was coming from the cover crops. Winter cover crop biomass alone explained 73% of the variations in microbial P and 33% of total labile P (Table 3).

Table 3: Spearman's rank correlation matrixes for winter cover crop biomass and maize stover input (summed over 4 years) versus phosphorus pools and maize plant P concentration

	Winter cover crop biomass	**Total biomass (maize stover + winter cover crop biomass)**	**P concentration of maize plants**
$NaHCO_3$-P_i	0.211 ns	0.26 ns	0.146 ns
HCl-P_i	0.288 ns	0.379*	0.8965***
$NaHCO_3$-P_o	0.175 ns	0.1 ns	0.181ns
Microbial-P	0.73**	0.528*	0.0155 ns
HCl-P_o	0.059 ns	-0.027 ns	0.0283 ns
Fulvic acid-P	0.042 ns	0.213 ns	0.1941 ns
Total labile P	0.33*	0.399*	0.581*
Humic acid-P	0.353*	0.1 ns	0.11 ns
H_2SO_4-P_o	-0.22 ns	-0.154 ns	-0.1352 ns
Total P	0.228 ns	0.251 ns	0.318 ns
P concentration of maize plants	0.327*	0.47*	–

*ns, not significant; *p<0.05; **p<0.01; ***p<0.001*

Under no-till, P tends to be stratified, which may partly explain these high values of $NaHCO_3$-P_i (>40 mg/kg, Figure 1) in the surface soil. It is noted that elevated levels of plant available P_i in the soil can induce Zn and Cu deficiency.[25] Further studies may therefore be needed to evaluate the effects of the winter cover crops on Zn and Cu availability.

Conclusions

The maize–winter cover rotations increased total P and some labile P pools in the surface soil when compared with the maize–fallow rotation, and this effect was positively correlated to cover crop biomass. The HCl-P_i pool was strongly correlated to maize seedling tissue P concentration and thus P supply for early maize growth. Non-application of fertiliser to maize, however, increased the accumulation of the recalcitrant humic-P fraction on the maize–oats rotation. Overall, the contribution from the winter cover crops to P availability in the surface soil suggests that, in the long term, fertiliser P could be reduced in low fertiliser input CA systems. Further work is recommended to evaluate the effects of winter cover crops on other soil nutrient pools such as Zn and Cu.

References

1. Cordell D, Drangert J, White S. The story of phosphorus, global food security and food for thought. Global Environ Chang. 2009;19:292–305. http://dx.doi.org/10.1016/j.gloenvcha.2008.10.009
2. Ward J. Peak phosphorus, quoted reserves vs. production history. Energy Bulletin [serial on the Internet]. 2008 Aug 26 [cited 2014 Feb 17]. Available from: http://www.resilience.org/stories/2008-08-26/peak-phosphorus-quoted-reserves-vs-production-history
3. Franzluebbers AJ. Introduction to special section – Supporting ecosystem services with conservation agricultural approaches. Renew Agr Food Syst. 2013;28:99–101. http://dx.doi.org/10.1017/S1742170513000021
4. Mannering JV, Griffith DR, Johnson KD. Winter cover crops – Their value and management. West Lafayette, IN: Department of Agronomy, Purdue University Cooperative Extension Service; 2007. Available from: http://www.agry.purdue.edu/ext/forages/rotational/articles/PDFs-pubs/winter-cover-crops.pdf
5. Gathumbi SM, Cadisch G, Buresh RJ, Giller KE. Subsoil nitrogen capture in mixed legume stands as assessed by deep nitrogen -15 placement. Soil Sci Soc Am. 2003;67:573–582. http://dx.doi.org/10.2136/sssaj2003.5730
6. Karasawa T, Kasahara Y, Takebe M. Differences in growth responses of maize to preceding cropping caused by fluctuation in the population of indigenous arbuscular mycorrhizal fungi. Soil Biol Biochem. 2002;34:851–857. http://dx.doi.org/10.1016/S0038-0717(02)00017-2
7. Dube E, Chiduza C, Muchaonyerwa P. Conservation agriculture effects on soil organic matter on a Haplic Cambisol after four years of maize-oat and maize-grazing vetch rotations in South Africa. Soil Till Res. 2012; 123:21–28. http://dx.doi.org/10.1016/j.still.2012.02.008
8. Murungu FS, Chiduza C, Muchaonyerwa P, Mnkeni PNS. Mulch effects on soil moisture and nitrogen, weed growth and irrigated maize productivity in a warm-temperate climate of South Africa. Soil Till Res. 2011;112:58–65. http://dx.doi.org/10.1016/j.still.2010.11.005
9. Maroko JB, Buresh RJ, Smithson PC. Soil phosphorus fractions in unfertilized fallow-maize systems on two tropical soils. Soil Sci Soc Am. 1999;63:320–326. http://dx.doi.org/10.2136/sssaj1999.03615995006300020009x
10. Hedley MJ, Stewart JWB, Chauhan BS. Changes in inorganic and organic soil phosphorus fractions induced by cultivation practices and by laboratory incubations. Soil Sci Soc Am. 1982;46:970–976.
11. Tate KR. Soil phosphorus. In: Vaughan D, Malcolm RE, editors. Soil organic matter and biological activity, developments in plant and soil sciences. Vol 16. Dordrecht: Martinus Nijhoff/Dr W. Junk Publishers; 1985. p. 329–377. http://dx.doi.org/10.1007/978-94-009-5105-1_10
12. Bowman RA, Cole CV. An exploratory method for fractionation of organic phosphorus from grassland soils. Soil Sci. 1978;125:95–101. http://dx.doi.org/10.1097/00010694-197802000-00006
13. Lupwayi NZ, Clayton GW, O'Donovan JT, Harker KN, Turkington TK, Soon YK. Phosphorus release during decomposition of crop residues under conventional and zero tillage. Soil Till Res. 2007;95:231–239. http://dx.doi.org/10.1016/j.still.2007.01.007
14. Brookes PC, Powlson DS, Jenkinson DS. Phosphorus in the soil microbial biomass. Soil Biol Biochem. 1984;16:169–175. http://dx.doi.org/10.1016/0038-0717(84)90108-1
15. Kovar JL, Pierzynski GM. Methods of phosphorus analysis for soils, sediments, residuals, and waters – Revised edition. Southern Cooperative Series Bulletin No 408. Blacksburg, VA: Virginia Tech University; 2009. Available from: http://www.sera17.ext.vt.edu/Documents/P_Methods2ndEdition2009.pdf
16. Thien SJ, Myers R. Determination of bioavailable phosphorus in soil. Soil Sci Soc Am. 1992;56:814–818. http://dx.doi.org/10.2136/sssaj1992.03615995005600030023x
17. Okalebo JR, Gathua KW, Woomer PL. Laboratory methods of soil and plant analysis, a working manual. Nairobi: TSBF-KARI, SSEA, SACRED Africa; 2002.
18. Murphy J, Riley JP. A modified single solution method for the determination of phosphate in natural waters. Anal Chim Acta. 1962;27:31–36. http://dx.doi.org/10.1016/S0003-2670(00)88444-5
19. Tiessen H, Moir JO. Characterization of available P by sequential extraction. In: Carter MR, editor. Soil sampling and methods of analysis. Boca Raton: CRC Press; 1993.
20. Pavinato PS, Merlin A, Rosolem CA. Phosphorus fractions in Brazilian Cerrado soils as affected by tillage. Soil Till Res. 2009;105:149–155. http://dx.doi.org/10.1016/j.still.2009.07.001
21. Agbenin JO, Tiessen H. Phosphorus forms in particle size fractions of a toposequence from Northeast Brazil. Soil Sci Soc Am. 1995;59:1687–1693. http://dx.doi.org/10.2136/sssaj1995.03615995005900060026x
22. Arlauskienė A, Maikštėnienė S, Šlepetienė A. Effect of cover crops and straw on the humic substances in the clay loam Cambisol. Agron Res. 2010;8:397–402.
23. Noack SR, McLaughlin MJ, Smernik RJ, McBeath TM, Armstrong RD. Crop residue phosphorus: Speciation and potential bio-availability. Plant Soil. 2012;359(1–2):375–385. http://dx.doi.org/10.1007/s11104-012-1216-5
24. Sharpley AN, Smith SJ. Fractionation of inorganic and organic phosphorus in virgin and cultivated soils. Soil Sci Soc Am. 1985;49:127–130. http://dx.doi.org/10.2136/sssaj1985.03615995004900010025x
25. Landon JR. Booker tropical soil manual: A handbook for soil survey and agricultural land evaluation in the tropics and subtropics. New York: Booker Agriculture International Ltd., Longman; 1984.

Flood variation and soil nutrient content in floodplain vegetation communities in the Okavango Delta

AUTHORS:
Gaolathe Tsheboeng[1]
Mphaphi Bonyongo[1]
Mike Murray-Hudson[1]

AFFILIATION:
[1]Okavango Research Institute, Maun, Botswana

CORRESPONDENCE TO:
Gaolathe Tsheboeng

EMAIL:
gtsheboeng@tati.ub.bw

POSTAL ADDRESS:
Okavango Research Institute, Private Bag 285, Maun 00000, Botswana

DATES:
Received: 31 May 2013
Revised: 25 Oct. 2013
Accepted: 30 Oct. 2013

KEYWORDS:
flood; communities; soil; vegetation; zones

HOW TO CITE:
Tsheboeng G, Bonyongo M, Murray-Hudson M. Flood variation and soil nutrient content in floodplain vegetation communities in the Okavango Delta. S Afr J Sci. 2014;110(3/4), Art. #2013-0168, 5 pages. http://dx.doi.org/10.1590/sajs.2014/20130168

We investigated the influence of hydroperiod variation on soil nutrient content in the Okavango Delta seasonal floodplains. Soil samples were collected from eight zones of homogenous vegetation cover after low and high floods and analysed for pH, Na, Mg, Ca, K and P content. A Student's *t*-test was used to test for differences in pH, Na, Mg, Ca, K and P between soils after low and high floods. The Kruskal-Wallis test was used to compare means of flooding duration and depth between low and high floods. Na, K, Mg, P and pH levels in soils were significantly different ($p<0.05$) after low floods compared with after high floods. Na content was lower ($p<0.05$) in Zones 2 and 8 and higher ($p<0.05$) in Zones 4, 5 and 6 during high flood than during low flood. Ca content was lower ($p<0.05$) in Zones 1, 2, 3, 4 and 5 and higher in Zone 7 ($p<0.05$) under high flood than during low flood. Mg content was lower ($p<0.05$) in Zones 1 and 5 and higher ($p<0.05$) in Zones 6, 7 and 8 during high flood than during low flood. K content was lower in Zones 1 and 3 and lower in Zones 6, 7 and 8 ($p<0.05$) during high flood than during low flood. pH was significantly lower and higher after a high flood in Zones 1 and 6, respectively, than during low flood. P content was significantly ($p<0.05$) higher in all zones after high flood than after low flood. Flooding depth and duration increased ($p<0.05$) in all vegetation zones during high flood. Our results have direct implications for *molapo* (flood recession) farming. We recommend that farmers plough immediately after the onset of flood recession when the soil is still moist and rich in nutrients.

Introduction

Soil nutrient dynamics in seasonal floodplain ecosystems are highly complex[1] as a result of flood pulses and changing redoximorphic state.[2,3] Flood pulse refers to the alternating dry and wet conditions in floodplain ecosystems.[4] It facilitates soil nutrient exchange between rivers and their associated seasonal floodplains.[5] During floods, soil nutrients dissolve in floodwaters and are transported from seasonal floodplain surfaces into adjacent rivers.[4] Soil nutrients may also be transported from the river into seasonal floodplains through lateral flow.[4] Flooding can lead to both increases and decreases in soil nutrient content. During flooding the soil becomes highly reduced, resulting in a decrease in pH which leads to an increase in the mobility of soil nutrients such as P, N, Mg, Ca, Na and K.[2] These nutrients include those that were deposited by the previous flood and those released from organic matter decomposition accumulated during dry periods.[6] Soil flooding can cause hypoxia leading to a reduction in the soil nutrient content available to plants.[7] As a result of hypoxia, the organic matter decomposition rate is reduced[3], leading to low soil nutrient content release[8].

Although the relationship between hydroperiod (flooding duration and depth) and soil nutrients has been extensively studied elsewhere,[1,9-12] the Okavango Delta has received little attention in this regard. Two studies have shown[13,14] that soil nutrient status in the Okavango Delta seasonal floodplains is influenced by hydroperiod. As a result of the soil moisture content gradient, seasonal floodplains experience unequal distribution of soil nutrients. Generally, soil nutrient content is higher in primary floodplains, followed by secondary floodplains, with tertiary floodplains having low soil nutrient content.[13,14] However, these studies[13,14] were conducted during a low flood (1996/1997 flooding season) in which flooding duration and depth were, respectively, short and shallow.

Since 2005, floods in the Okavango Delta have increased annually, with the highest floods to date recorded in 2010.[15] The 1996/1997 low flood (389.33 m^3/s) season measured at Mohembo was preceded by five consecutive years of low flood while the 2010 high flood (1043.33 m^3/s) was preceded by five consecutive years of high floods.[15] During high flood, flooding duration is prolonged and flooding depth increases. The surface area inundated (flood extent) also increases and as a result some tertiary floodplains which are usually not flooded during low floods become inundated. It is still not known how soil nutrient content in the Okavango Delta seasonal floodplains is influenced by a high flood. The Okavango Delta seasonal floodplains are important for the local people as they use them for crop production (*molapo* farming or flood recession farming). Because flooding is an important factor in nutrient cycling in the Delta, there is a need to establish how variation in hydroperiod affects its soil nutrient dynamics. Soil nutrient availability is significant as it supports primary production in seasonal floodplain vegetation communities.[13] We therefore compared soil nutrient content after low (1996/1997) and high (2010) flood conditions. It was hypothesised that soil nutrient content would be lower during high flood than during low flood conditions.

Materials and methods

Study area

The study was conducted in the Nxaraga seasonal floodplains in the Okavango Delta (Figure 1). Nxaraga seasonal floodplains receive floods between May and October with approximately 500 mm of rainfall per year.[16] The mean maximum summer temperature is 30.5 °C with a mean minimum of 14.8 °C.[17] During winter, mean maximum temperature ranges between 25.3 °C and 28.7 °C while mean minimum temperature ranges between 7.0 °C and 10.0 °C.[17]

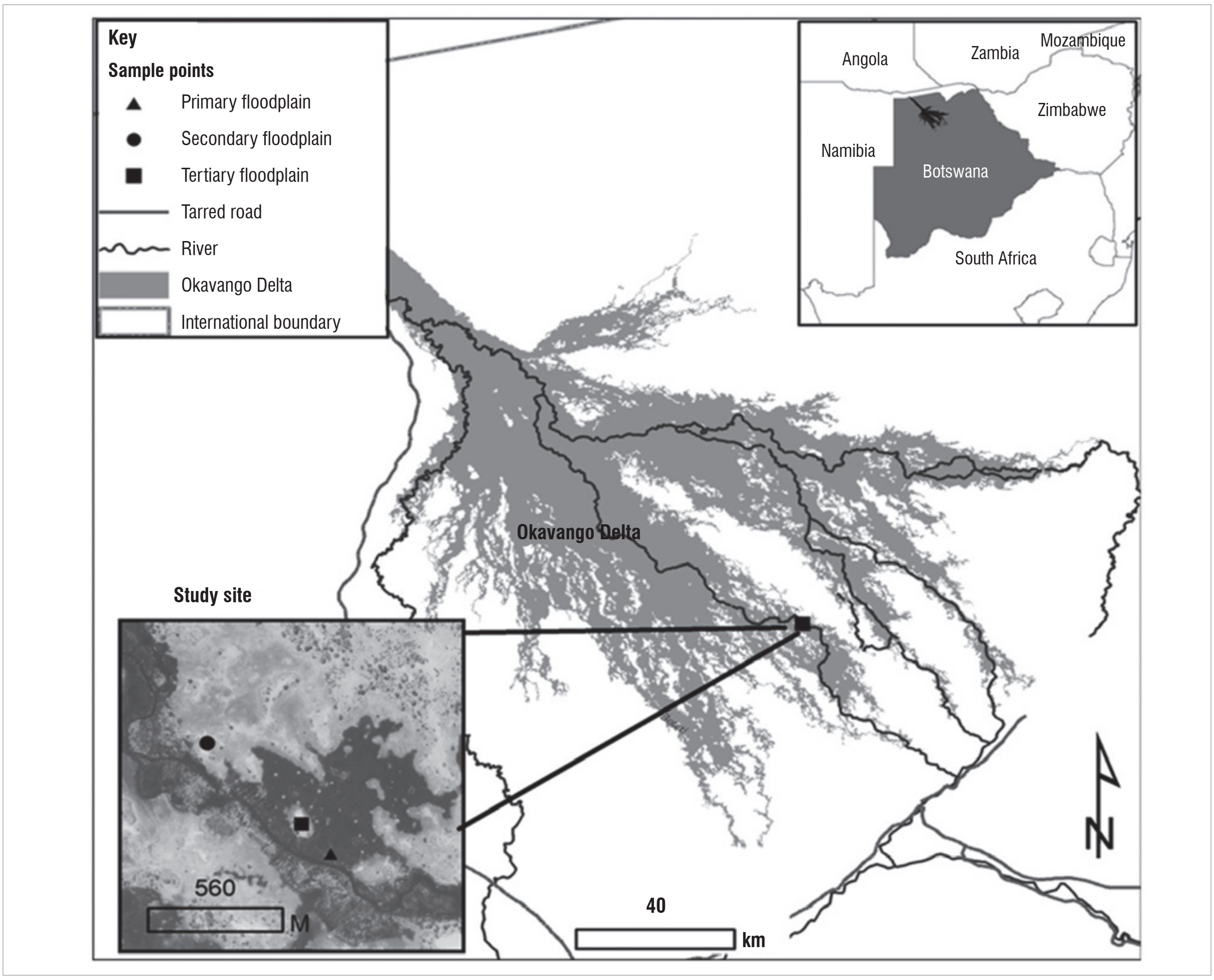


Figure 1: The Okavango Delta showing the location of the study area and floodplains (bottom left corner).

Soil sampling

The comparison of soil nutrient content between years 2010 and 1997 was done through re-sampling soil from zones sampled during low flood[13] selected on the basis of homogenous vegetation cover along an elevation gradient between a river channel (low) and dry land (high). During low flood the vegetation zones were dominated by *Alternanthera sessilis* and *Ludwigia stolonifera* (Zone 1), *Cyperus articulatus* and *Schoenoplectus corymbosus* (Zone 2), *Miscanthus junceus* and *Digitaria scalarum* (Zone 3), *Paspalidium obtusifolium* and *Panicum repens* (Zone 4), *Setaria sphacelata* and *Eragrostis inamoena* (Zone 5), *Imperata cylindrica* and *Setaria sphacelata* (Zone 6), *Vetiveria nigritana* and *Setaria sphacelata* (Zone 7) and *Sporobolus spicatus* and *Cynodon dactylon* (Zone 8).[18] During high flood, co-dominant species changed into *Oxycaryum cubense* and *Vossia cuspidata* (Zone 1), *Cyperus articulatus* and *Schoenoplectus corymbosus* (Zone 2), *Miscanthus junceus* and *Pycreus flavescens* (Zone 3), *Eleocharis dulcis* and *Leersia hexandra* (Zone 4), *Eragrostis inamoena* and *Panicum repens* (Zone 5), *Imperata cylindrica* (Zone 6), *Vetiveria nigritana* and *Setaria sphacelata* (Zone 7), and *Sporobolus spicatus* and *Sporobolus acinifolius* (Zone 8).[19] Zones 1 and 2 were located in the primary floodplains which are closer to the river channel while Zones 3, 4, 5, 6 and 7 were located in secondary floodplains on relatively intermediately elevated areas. Zone 8 was found in the tertiary floodplains, which are relatively highly elevated areas.[13]

A total of 40 plots from an earlier study[13] were re-sampled in September 2010. Soil samples were collected from five 5 m x 5 m plots in each of the eight pre-defined zones. From each plot, one soil sample was collected at a depth of between 0 and 300 mm using a soil auger. The soil samples were oven dried at 80 °C for 24 h, ground using a pestle and mortar and sieved through a 2-mm mesh. The samples were analysed for pH, extractable Na, P, Ca, Mg and K. pH was determined from a 1:1 soil: water suspension using a pH electrode (model 330i, Wissenschaftlich-Technische Werkstatten GmbH, Weilheim, Germany). K and Na content was determined using a flame photometer (model 410, Sherwood Scientific, Cambridge, UK) while Mg and Ca content was determined using a Varian atomic absorption spectrophotometer (Model AA 220, Varian, Sydney, Australia). Total P analysis was performed using a Bran+Luebbe Auto Analyser 3 (Norderstedt, Germany).

Hydroperiod sampling

Flooding depth was determined using a calibrated 2-m polyvinyl chloride pipe in all the plots where the soil samples were collected. Flooding duration was estimated as the number of weeks in which the plots were flooded.

Statistical data analysis

A Student's *t*-test was performed to compare soil nutrient content between low and high flood conditions in different floodplain vegetation zones. The Kruskal-Wallis test was used to compare the average flooding duration and depth between a low and high flood. Correlations between flooding depth, duration of flooding and soil nutrient status were performed using a Spearman's rank correlation with SPPS version 19, 2010.

Results

Flooding depth and duration were significantly ($p<0.05$) higher in all vegetation zones during high flood than during low flood (Table 1).

Table 1: Flooding depth and flooding duration during low (1996/1997) and high (2010) floods in Nxaraga, Okavango Delta

	Flooding depth		Flooding duration	
Zone	**Low flood**	**High flood**	**Low flood**	**High flood**
1	0.5	1.62*	23	P*
2	0.78	1.18*	16	P*
3	0.48	0.72*	12	P*
4	0.35	0.83*	9	P*
5	0.15	0.6*	5	P*
6	0	0.5*	0	12*
7	0	0.32*	0	6*
8	0	0.12*	0	3*

*P, permanently flooded; *p<0.05*

Ca content was lower ($p<0.05$) in Zones 1, 2, 3, 4 and 5 and higher in Zone 7 ($p<0.05$) under high flood conditions in 2010 than under low flood conditions in 1997 (Figure 2). Other differences (Zone 6) were not found to be significant at the $p<0.05$ level. Na content was lower ($p<0.05$) in vegetation Zones 2 and 8 and higher ($p<0.05$) in Zones 4, 5 and 6 during high flood than during low flood (Figure 3). Mg content was lower ($p<0.05$) in vegetation Zones 1 and 5 and higher ($p<0.05$) in Zones 6, 7, 8 during high flood than during low flood (Figure 4). K content was lower in vegetation Zones 1 and 3 and higher in Zones 6, 7 and 8 ($p<0.05$) during high flood than during low flood (Figure 5).

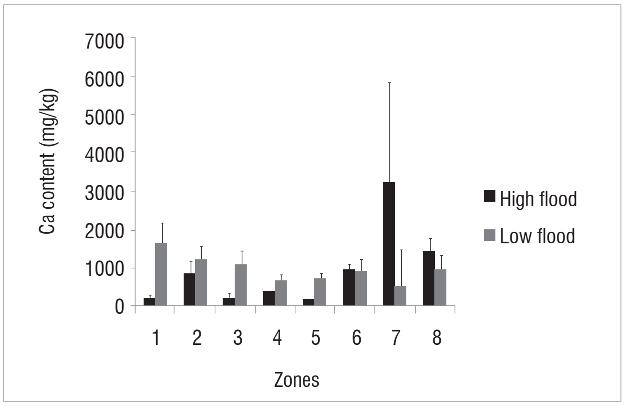


Figure 2: Mean Ca content after low and high floods.

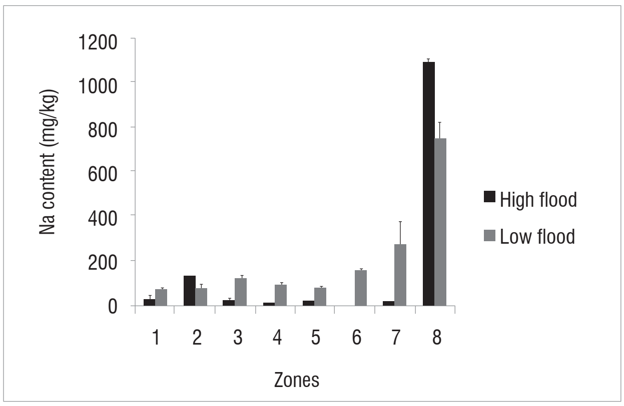


Figure 3: Mean Na content after low and high floods.

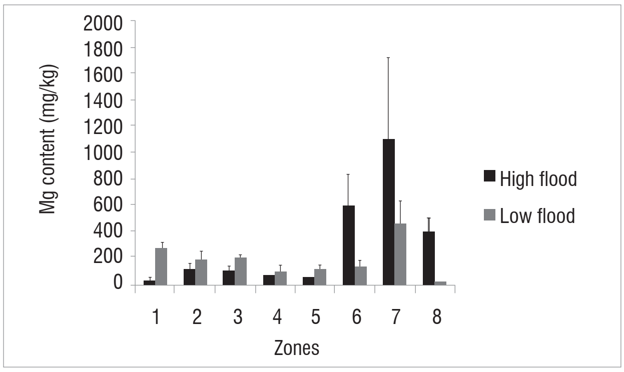


Figure 4: Mean Mg content after low and high floods.

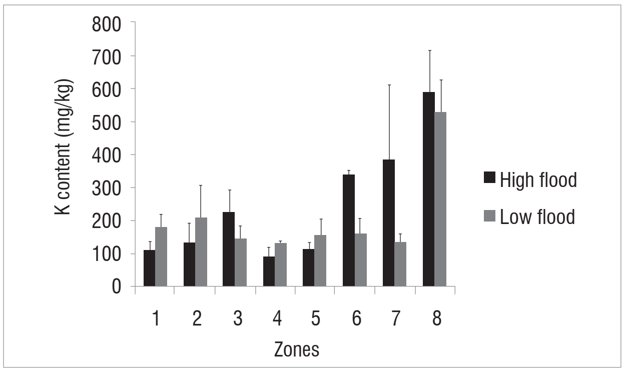


Figure 5: Mean K content after low and high floods.

pH was significantly ($p<0.05$) higher in Zone 1 and lower in Zone 6 during high flood than during low flood (Figure 6).

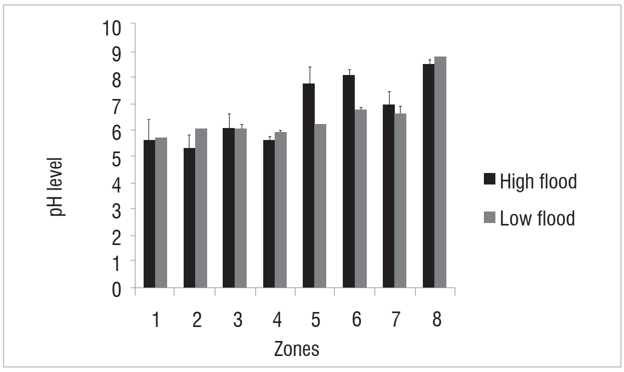


Figure 6: Mean pH after low and high floods.

P content was significantly ($p<0.05$) higher in all Zones after the high flood than after the low flood (Figure 7).

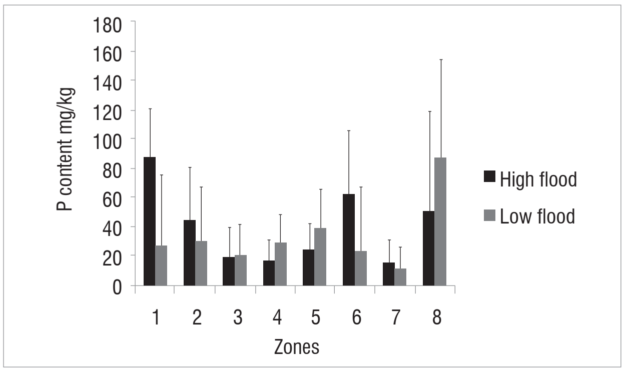


Figure 7: Mean P content after low and high floods.

Correlation between hydroperiod and soil nutrient content

In 2010, Na, Ca, K and Mg content and pH level were negatively correlated with flooding depth and duration. P content was positively correlated with flooding depth and duration (Table 2).

Table 2: Correlations (*r*-values) between flooding depth, duration of flooding and soil nutrient content and pH level

Soil nutrient	Flooding depth	Flooding duration
Na	-0.43	-0.49
Ca	-0.47	-0.54
K	-0.40	-0.40
Mg	-0.59	-0.70
P	0.60	0.52
pH	-0.50	-0.48

Discussion

Soil nutrient content varied after low and high flood. With the exception of P, soil nutrient content generally increased with decreasing flooding depth and duration. K, Ca, Mg and Na contents were lower in Zones 1, 2 and 3 and higher in Zones 6, 7 and 8 in response to a high flood. These results are consistent with findings from previous studies.[1,11,12,20,21] High K content in zones experiencing low flooding depth and short flooding duration was also observed in the seasonal floodplains of the Orinoco River[1] and the Rhine and Narow Rivers[12]. In a study conducted in the Careiro Islands in the central Amazon, Na content was higher in dry zones than in flooded zones.[20] The findings of this study also agree with the results from a study[21] conducted in a Mongolian lake in which Ca content decreased with an increase in moisture content. It was also observed that Ca content increased with decreasing moisture content in a study conducted on the Lower Mkuze seasonal floodplain in South Africa.[12]

Low Ca, Mg, K and Na contents in vegetation Zones 1, 2, 3, 4 and 5 after a high flood could be a result of leaching and dilution[22-24] because flooding increases the solubility of mineral nutrients[2]. It could be expected that during a high flood more soil nutrients dissolve in water and are lost through leaching as water infiltrates the soil. It could also be expected that because clay is negatively charged[18] cations would bond to the soil particles, thus reducing leaching. However, leaching of cations has been found to be accelerated by dissociation of NO_3^- from HNO_3 (from nitrification).[18] Zones 1, 2, 3, 4 and 5 experienced high flooding depth and long flooding duration, which would suggest that more of their soil nutrients dissolved in the water and were lost through leaching.

Another factor that could lead to reduced soil nutrients during high flood is the rate of decomposition of organic matter. Organic matter is a reservoir of nutrients which are released when it decomposes.[25] During flooding, water displaces oxygen from the soil,[26,27] leading to anaerobic conditions.[28-30] Under anaerobic conditions, the rate of decomposition of organic matter declines, resulting in low soil nutrient content.[3] It is likely that Zones 1, 2, 3, 4 and 5 experienced anaerobic conditions because of their relatively high flooding depth and long flooding duration, resulting in a low organic matter decomposition rate, and hence low soil nutrient content. In contrast K, Mg, Na and pH increased in vegetation Zones 6, 7 and 8 under high flood conditions.

High K, Mg, Na and pH after high floods in Zones 6, 7 and 8 could be attributed to increased organic matter decomposition rates, evapotranspiration and lateral flow deposition. Sediment deposition in floodplains leads to an increase in soil nutrients.[29] Water and sediments from the main river channel are a source of dissolved nutrients to the floodplains.[4,31] During high flood, vegetation Zones 1, 2, 3 and 4 were almost converted into permanent swamps and as a result soil nutrients could have been transported away from them into the peripheral Zones 6, 7 and 8 through lateral flow.[31] Increased soil nutrients in Zones 6, 7 and 8 could also be attributed to organic matter accumulation and decomposition. During low floods, these zones were not inundated[14] and as a result organic matter may have accumulated.[31] When they received water during high floods, organic matter decomposition may have been triggered and nutrients released.[6] Furthermore, when organic matter dries, it mineralises into soil nutrients such as Ca, thus increasing the soil content of these cations.[32]

P content was higher after a high flood than after a low flood in all vegetation zones. P has a strong affinity to fine clay particles and it could be expected that during high floods more sediment is deposited, consequently leading to high P content.[2,3,33] The high affinity of P (which exists as PO_4^{-3} in solution) to clay is a result of lanthanum (has either +2 or +3 charge) that is embedded in the clay structure.[34] P binding to clay is also influenced by pH. It was found that P binding to clay decreased with an increase in pH from 7 to 9, which was attributed to formation of hydroxyl species of the lanthanum ions decreasing the number of P binding sites on the clay sites.[34] A high flood is expected to lead to anoxic conditions because of increased water depth and prolonged waterlogging, which leads to mobilisation of P and results in its increase.[2,3] Under aerobic conditions, P binds to iron oxides. Because of prolonged anaerobic conditions imposed by flooding, Fe bound to P is reduced from Fe (III) to Fe (II), releasing P from iron-phosphate complexes.[2,3,35] It is expected that during low flooding conditions, P reacts with Ca, Al and Fe oxyhydroxides as a result of aerobic conditions, consequently reducing its available content in the soil.

Conclusion

Soil nutrient content increased in some vegetation zones while it decreased in others after high flooding. Ca and Mg content increased in Zone 7. Mg content also increased in Zones 6 and 8. K content increased in Zones 6, 7 and 8 while P content increased in all the zones. Na content increased in all zones except Zones 1 and 8. These results therefore show that flooding variation influences soil nutrient content in the Okavango Delta seasonal floodplains. These results have direct implications for *molapo* (flood recession) farming. We recommend that farmers plough immediately after the onset of flood recession when the soil is still moist and rich in nutrients.

Authors' contributions

G.T. was the project leader and was responsible for the data collection, analysis and write up. M.B. was responsible for the data collection and for providing technical writing advice. M.M-H. assisted in the statistical data analysis and provided technical writing advice.

References

1. Dezzeo N, Herrera R, Escalanta G, Chacòn. Deposition of sediments during a flooding event on seasonally flooded area of the lower Orinoco River and two of its black water tributaries, Venezuela. Biogeochemistry. 2000;49:241–257. http://dx.doi.org/10.1023/A:1006375101368
2. Mitsch WJ, Gosselink JG. Wetlands. New York: John Wiley and Sons; 2000.
3. Gallardo A. Spatial variability of soil properties in a floodplain forest in Northwest Spain. Ecosystems. 2003;6:564–576. http://dx.doi.org/10.1007/s10021-003-0198-9
4. Junk WJ, Bayley PB, Sparks RE. The flood pulse concept in river-floodplain systems. In: Dodge DP, editor. Proceedings of the International Large River Symposium. Canadian special publication of fisheries and aquatic sciences. Ottawa: Department of Fisheries and Oceans; 1989. p. 110–127.
5. Valett HM, Baker MA, Morrice JA, Crawford CS, Molles Jr MC, Dahm CN, et al. Biogeochemical and metabolic responses to the flood pulse in a semiarid floodplain. Ecology. 2005;86(1):220–234. http://dx.doi.org/10.1890/03-4091
6. Powell WG. Identifying land use/land cover (LULC) using data as a hydrologic model input for local floodplain management [research report]. San Marcos,TX: Texas State University; 2009.

7. Chen H, Qualls RG, Blank RR. Effect of soil flooding on photosynthesis, carbohydrate partitioning and nutrient uptake in the invasive exotic *Lepidium latifolium*. Aquat Bot. 2005;82:250–268. http://dx.doi.org/10.1016/j.aquabot.2005.02.013

8. Teyler G, Olsson T. Concentration of 60 elements in the soil solution as related to the soil acidity. Euro J S Sci. 2001;52:151–165. http://dx.doi.org/10.1046/j.1365-2389.2001.t01-1-00360.x

9. Terrill TH, Allen VG, Cranford JP, Foster JG. Influence of moisture on chemical composition of annual ryegrass and digestibility by meadow voles. Virg J Sci. 1991;42(1):101–111.

10. Yavitt JB, Wieder RK, Wright JS. Soil nutrient dynamics in response to irrigation of a Panamanian tropical moist forest. Biogeochemistry. 1993;19(1):1–25. http://dx.doi.org/10.1007/BF00000572

11. Antheunise AM, Verhoeven JT. Short-term responses of soil nutrient dynamics and herbaceous riverine plant communities to summer inundation. Wetlands. 2008;28(1):232–244. http://dx.doi.org/10.1672/06-33.1

12. Humphries M. Sedimentation and chemical processes on the Lower Mkuze floodplain: Implications for wetland structure and function [PhD thesis]. Durban: University of KwaZulu-Natal; 2008.

13. Bonyongo MC. Vegetation ecology of the seasonal floodplains in the Okavango Delta, Botswana [MSc thesis]. Pretoria: University of Pretoria; 1999.

14. Bonyongo MC, Mubyana T. Soil nutrient status in vegetation communities of the Okavango Delta floodplains. S Afr J Sci. 2004;100:337–340.

15. Okavango Research Institute. Okavango Delta monitoring and forecasting [homepage on the Internet]. c2011 [cited 2011 Mar 23]. Available from: http://okavangodata.ub.bw/ori/monitoring/water

16. Wilson BH, Dincer T. An introduction to the hydrology and hydrography of the Okavango Delta. Proceedings of the Symposium on the Okavango Delta and its Future Utilization; Gaborone: National Museum; 1976.

17. Ellery K, Ellery WN, Rogers KH, Walker BH. Water depth and biotic insulation: Major determinants of back-swamp plant community composition. Wetlands Ecol Manage. 1991;1:149–162. http://dx.doi.org/10.1007/BF00177289

18. Barber SA. Soil nutrient bioavailability: A mechanistic approach. New York: Library of Congress; 1995.

19. Tsheboeng G, Bonyongo MC, Murray-Hudson M. Influence of flood variation on seasonal floodplain vegetation communities in the Okavango Delta. J Aqua Sci. In press 2014.

20. Alfaia SS, Falcao NP. Study of nutrient dynamics in the floodplain soils of Careiro Island-Central Amazonia. Amazon. 1993;12(3–4):485–493.

21. Strauss A, Schickoff U. Influence of soil-ecological conditions on vegetation zonation in a western Mongolian Lake shore-semi-desert ecotone. Erdkunde. 2007;61(1):72–92. http://dx.doi.org/10.3112/erdkunde.2007.01.05

22. Conklin AR. Introduction to soil chemistry: Analysis and instrumentation. New Jersey: John Wiley and Sons; 2005. http://dx.doi.org/10.1002/0471728225

23. Day FP. Litter decomposition rates in the seasonally flooded great dismal swamp. Ecology. 1982;63:670–678. http://dx.doi.org/10.2307/1936787

24. De Datta SK. Principles and practices of rice production. New York: John Wiley and Sons; 1981.

25. Gosz JR. Organic matter and nutrient dynamics of the forest floor in the Hubbard Brook forest. Oecologia. 1976;22:305–320. http://dx.doi.org/10.1007/BF00345310

26. Lobo PC Jolly CA. Tolerance of hypoxia and anoxia in neotropical tree species. In: Scarano FR, Franco AC, editors. Ecophysiological strategies of xerophytic and amphibious plants in the neotropics. Oecol Bras. 1998;5:137–156.

27. Hefting M, Clement JC, Dowrick D, Losanday A-C, Bernal S, Cimpian C, et al. Water table elevation controls on soil nitrogen cycling in riparian wetlands along European climatic gradient. Biogeochemistry. 2004;67:113–134. http://dx.doi.org/10.1023/B:BIOG.0000015320.69868.33

28. Kozlowski TT. Plant responses to flooding of soil. Bioscience. 1984;34:162–167. http://dx.doi.org/10.2307/1309751

29. Moorhead KK, McAuthur J. Spatial and temporal patterns of nutrient concentrations in foliage of riparian species. Am Mid Nat. 1996;136:29–41. http://dx.doi.org/10.2307/2426629

30. Kozlowski TT. Responses of woody plants to flooding and salinity. Tree Physiol Monogr [serial on the Internet]. 1997[cited 2011 Feb 10];1. Available from: http://www.pucrs.br/fabio/fisiovegetal/Encharcamento.pdf

31. Junk WJ. General aspects of floodplain ecology with special reference to Amazonian floodplains: In: Junk WJ, editor. The central Amazon floodplain: Ecology of a pulsing system. New York: Springer-Verlag; 1997. p. 3–20. http://dx.doi.org/10.1007/978-3-662-03416-3_1

32. Kalisz B, Łachacz A. Content of nutrients, heavy metals and exchangeable cations in riverine organic soils. Pol J S Sci. 2009;42:43–52.

33. Harry OV, Verman E, Mario P, Frank W, Lee VD, Guda EM, et al. Importance of sediment deposition and denitrification for nutrient retention in floodplain wetlands. Appl Veg Sci. 2006;9:163–174. http://dx.doi.org/10.1111/j.1654-109X.2006.tb00665.x

34. Ross G, Haghseresht F, Cloete TE. The effect of pH and anoxic on the performance of Phoslock, a phosphorous binding clay. Harmful Algae. 2008;7:545–550. http://dx.doi.org/10.1016/j.hal.2007.12.007

35. Brady NC, Weil RR. The nature and properties of soils. Columbus, OH: Pearson; 2008.

Impact of lower stratospheric ozone on seasonal prediction systems

AUTHORS:
Kelebogile Mathole[1]
Thando Ndarana[1]
Asmerom Beraki[1]
Willem A. Landman[2,3]

AFFILIATIONS:
[1]South African Weather Service – Research, Pretoria, South Africa
[2]Council for Scientific and Industrial Research – Natural Resources and the Environment, Pretoria, South Africa
[3]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Kelebogile Mathole

EMAIL:
kelebogile.mathole@weathersa.co.za

POSTAL ADDRESS:
South African Weather Service – Research, South Bolepi House, 442 Rigel Avenue, Erasmusrand 0181, South Africa

DATES:
Received: 28 May 2013
Revised: 22 Oct. 2013
Accepted: 30 Oct. 2013

KEYWORDS:
polar vortex; eddy-driven jet; stratosphere; ozone depletion; ENSO

HOW TO CITE:
Mathole K, Ndarana T, Beraki A, Landman W. Impact of lower stratospheric ozone on seasonal prediction systems. S Afr J Sci. 2014;110(3/4), Art. #2013-0161, 8 pages. http://dx.doi.org/10.1590/sajs.2014/20130161

We conducted a comparison of trends in lower stratospheric temperatures and summer zonal wind fields based on 27 years of reanalysis data and output from hindcast simulations using a coupled ocean-atmospheric general circulation model (OAGCM). Lower stratospheric ozone in the OAGCM was relaxed to the observed climatology and increasing greenhouse gas concentrations were neglected. In the reanalysis, lower stratospheric ozone fields were better represented than in the OAGCM. The spring lower stratospheric/upper tropospheric cooling in the polar cap observed in the reanalysis, which is caused by a direct ozone depletion in the past two decades and is in agreement with previous studies, did not appear in the OAGCM. The corresponding summer tropospheric response also differed between data sets. In the reanalysis, a statistically significant poleward trend of the summer jet position was found, whereas no such trend was found in the OAGCM. Furthermore, the jet position in the reanalysis exhibited larger interannual variability than that in the OAGCM. We conclude that these differences are caused by the absence of long-term lower stratospheric ozone changes in the OAGCM. Improper representation or non-inclusion of such ozone variability in a prediction model could adversely affect the accuracy of the predictability of summer rainfall forecasts over South Africa.

Introduction

The El Nino Southern Oscillation (ENSO) phenomenon is the single biggest contributing factor to climate variability because of its large global impact.[1] Its effect on seasonal summer rainfall over South Africa is well documented.[2-4] Studies have shown that ENSO signals are usually associated with rainfall anomalies over the country, that is, above normal rainfall conditions are often associated with La Nina events and below normal rainfall with El Nino events. ENSO-forced predictability becomes even more enhanced during the austral summer as a result of tropical circulation that becomes dominant during this season and thus increases the predictability of rainfall at shorter lead time scales.[5] Moreover, numerous modelling studies[5,6] have shown that variations of sea surface temperatures from the equatorial Pacific and Indian Oceans provide skillful predictions over southern Africa because of the linear relationship that they have with the region's summer seasonal rainfall.[7] Therefore, ENSO serves as a source for seasonal predictability over southern Africa, particularly in the case of above normal summer rainfall during La Nina years.[8]

Although ENSO-based seasonal prediction systems have come a long way to produce skillful summer rainfall forecasts during La Nina and El Nino events, they are constrained during neutral conditions over the equatorial Pacific Ocean as their skill diminishes.[9] Furthermore, ENSO explains only about 20–30% of the climate variability[10] over southern Africa. Stratosphere/troposphere coupling and stratospheric dynamics could therefore be explored and added as sources of seasonal predictability for the region; this notion is explored in this paper.

The eddy-driven jet, which dominates the circulation over the southern hemisphere during the summer,[11] and the associated storm tracks affect summer rainfall over South Africa.[12] The mechanism which is responsible for this association is explained by low-level baroclinicity.[13] An anomalously poleward position of the jet and storm tracks is associated with anomalously wet conditions over South Africa and an anomalously equatorward position is associated with anomalously dry conditions. During anomalously dry conditions, the cloud bands that bring much of the country's summer rainfall are displaced from their usual position and are located east of the country. Because the position of the jet is influenced by the strength of the polar vortex[14,15] through robust stratospheric and tropospheric coupling mechanisms, the variability of winter and spring stratospheric winds, as well as temperatures, could be a source of summer rainfall predictability.[16] Moreover, Son et al.[17] have shown that for stratospheric variability to be useful in predicting tropospheric processes, the former has to be represented correctly in a model.

At longer timescales, observational[18] and modelling studies[17,19] have shown that the formation of the ozone hole has led to lower stratospheric and upper tropospheric cooling during the austral spring months. This formation has also been responsible for the persistent poleward movement of the eddy-driven jet during the summer and a persistent positive phase of the Southern Annular Mode.[20] As would be expected, these changes in the tropospheric circulations have been accompanied by long-term changes in subtropical rainfall patterns.[21,22] However, these changes have not been caused by ozone depletion alone. Increasing greenhouse gas (GHG) concentrations have a cooling effect on the lower stratosphere.[23] During the 1970s to 2000, these two radiative forcings complemented each other.[24]

Because summer rainfall over South Africa is influenced by the interannual variability of the position of the jet,[12] it is reasonable to hypothesise that a model that incorrectly simulates the jet position variability and climatology will likely be unable to simulate rainfall variability correctly. The effect may subsequently compromise the reliability of rainfall predictions at the seasonal timescale, but improved representation of stratospheric processes – such as ozone depletion – in climate models used to predict climate variability might lead to improved seasonal forecasts.

The representation of stratospheric processes in climate models has various facets and is important for realistic simulations. A recent study[25] showed that if ozone variations in a model are zonally symmetrical as opposed to three dimensional, lower stratospheric and upper tropospheric temperature trends as well as changes in the zonal winds are underestimated. The proper representation of stratospheric ozone is achieved through the use of interactive stratospheric chemistry schemes such as McLandress et al.'s[19]. It is also possible that the atmospheric level at which the model top is located plays an important role in the accuracy of climate models in simulating stratospheric dynamics. The highest level in most models is 10 hPa, which is far too low to accurately capture the stratospheric polar vortex.

Information on stratospheric processes that occur at levels higher than 10 hPa is conveyed into the model vertical domain by specifying model top boundary conditions.[26] However, these boundary conditions may not be equivalent to actually including the stratospheric processes, which can be achieved by raising the model top to 0.01 hPa. An idealised modelling study[27] showed that stratospheric/tropospheric coupling is captured clearly in a model with the top as high as 0.1 hPa. High top models are also considered in CMIP5.[28] Increasing stratospheric resolution, in addition to the above, has the ability to improve seasonal climate predictions significantly.[29] Studies such as Roff et al.'s[30] also indicate the importance of stratospheric resolution on extended forecasting skill. All these issues are applicable at timescales longer than that of seasonal prediction but are relevant to this timescale and therefore raise many questions with regard to the role of stratospheric processes and seasonal predictability. As such, in this paper we consider the behaviour of lower stratospheric and upper tropospheric temperatures, as well as that of the eddy-driven jet of a coupled ocean-atmosphere general circulation model (OAGCM) in which the ozone representation is not realistic and has no GHG forcing. We then compare this behaviour to reanalysis data. This effort is to highlight the implications of forcing a seasonal prediction model with climatological stratospheric ozone fields that are zonally averaged.

Data and methods

We obtained hindcasts from the South African version of the coupled European Centre Hamburg Model (version 4.5) – Modular Ocean Model, version 3–South Africa, Ocean Atmosphere General Circulation Model (called the ECHAM 4.5-MOM3-SA OAGCM)[31] integrations for the first lead time (i.e. forecasts are made in early November for December-January-February).This model currently is used for operational forecast production at the South African Weather Service. Daily averages of zonal wind velocity and temperature fields over a period of 27 years (1983–2009) were constructed for the analyses. The coupled model output is available at a T42 (triangular truncation at wave number 42) horizontal resolution corresponding to a grid with 64 latitudes and 128 longitudes and with 19 vertical levels. The ozone field in the coupled model is relaxed toward the observed climatology and the anthropogenic forcing is neglected. The data set used here as a proxy for observation is from the National Centre for Environmental Prediction (NCEP) of the Department of Energy Reanalysis II[32] which is an updated version of the original NCEP data set.[33] Fixed fields such as lower stratospheric ozone and carbon dioxide (CO_2) concentrations have been improved. Seasonal climatology ozone is used in the radiation calculations to better represent processes associated with it. Therefore ozone depletion and increasing CO_2 are both presented more realistically than previously. NCEP reanalysis has a typical horizontal resolution of 73 latitude grids and 144 longitude grids (2.5º x 2.5º) with 17 vertical levels.

Using the two data sets (coupled model hindcasts and NCEP reanalysis), we investigated the hypothesis that lower stratospheric cooling is influenced by short-term variations and depletion in ozone. Linear trends were calculated by fitting least squares regression curves onto both data sets and then comparing them. The jet location was obtained by first calculating the zonal average of the zonal wind fields and then fitting cubic splines at all pressure levels followed by identifying the maximum value of the zonal wind over the whole of the troposphere. This approach bypasses problems associated with variations in the level at which the maximum zonal wind value occurs.

Climatology of the zonal wind

Figure 1a and 1b show the climatological general structure of the zonal wind flow during austral summer (DJF) as a function of latitude and pressure for both observations and coupled model hindcasts, respectively. This structure is characterised by positive westerlies covering the tropospheric region (within the 850 hPa and 10 hPa levels), with a strong wind maximum that is associated with the eddy-driven jet. The jet core in the observations and in the model occur at different levels. It is lower, located at 350 hPa, in the former but is higher than 350 hPa in the OAGCM. These zonal wind structures are caused by the strong meridional temperature gradient found in the middle latitudes, as required by the thermal wind balance.[34] The meridional temperature gradient is in turn caused by differential heating between the tropical and polar regions. Eddy momentum fluxes that converge in the middle latitudes[35] are responsible for maintaining the jet after having been transported poleward by anti-cyclonically breaking upper tropospheric troughs[34,36,37]. Also at play are energy conversions – barotropic processes that are associated with breaking waves convert eddy kinetic energy to mean kinetic energy.[34]

There also are significant differences between the NCEP and OAGCM jet structures (Figure 1a and 1b). The former shows a weaker jet core which is centred more poleward (at about 48˚S) than its OAGCM counterpart (at about 42˚S). However, the jet in the model appears to be more elevated than in the reanalysis, which could also have implications for moisture transport. As was alluded to in the introduction and will be discussed further below, the climatological position of the jet should be an important consideration in seasonal prediction systems because it determines the short-term (interannual) variability of the jet's position relative to South Africa. The position of the summer jet in the different data sets also indicates that the storm tracks would be placed at different locations in the observations and in the model. The association between the eddy-driven jet and storm tracks occurs through baroclinic waves which influence the zonal mean flow and hence storm tracks activity.[38] Storm tracks are important because they transport heat, moisture and momentum.[39]

Figure 1c and 1d give an indication of a relative climatological stratospheric wind circulation during austral winter (JJA) and spring (SON), respectively. These figures show zonal winds at 10 hPa, a level which is considered to be in the lower stratosphere as it is above the 500 K isentropic level for both data sets. Various studies[40,41] have used potential vorticity at this isentropic level to diagnose the dynamics of the polar vortex. A typical vortex comprises strong winter westerly winds replacing the summer and autumn easterlies.[42] The vortex decays in late spring and the timing of its break up is highly variable.[40,42]

The peaks of the winds at 10 hPa that occur in the mid-latitudes in Figure 1c and 1d are a good indication of the lower stratospheric polar night jet [43] in the data sets considered here[44], hence the associated polar vortex. NCEP climatological winds at 10 hPa are stronger than in the OAGCM case during the winter and spring. As noted above, a persistent stronger (or weaker) polar vortex during the winter and, in particular, spring, leads to a poleward (or equatorward) eddy-driven jet during summer.[14,15] Therefore, because the NCEP vortex is stronger than its OAGCM counterpart, one would expect the summer eddy-driven jet of the former to be positioned more poleward than that of the OAGCM. This displacement is indeed found to be the case. This effect is a purely dynamical phenomenon as it has been demonstrated by idealised modelling studies.[14] Idealised models comprise only dynamical cores and other physical processes[44] to separate atmospheric dynamical phenomena from other processes. The above discussion suggests that the climatology of stratospheric zonal winds is consistent with the tropospheric circulation in both data sets.

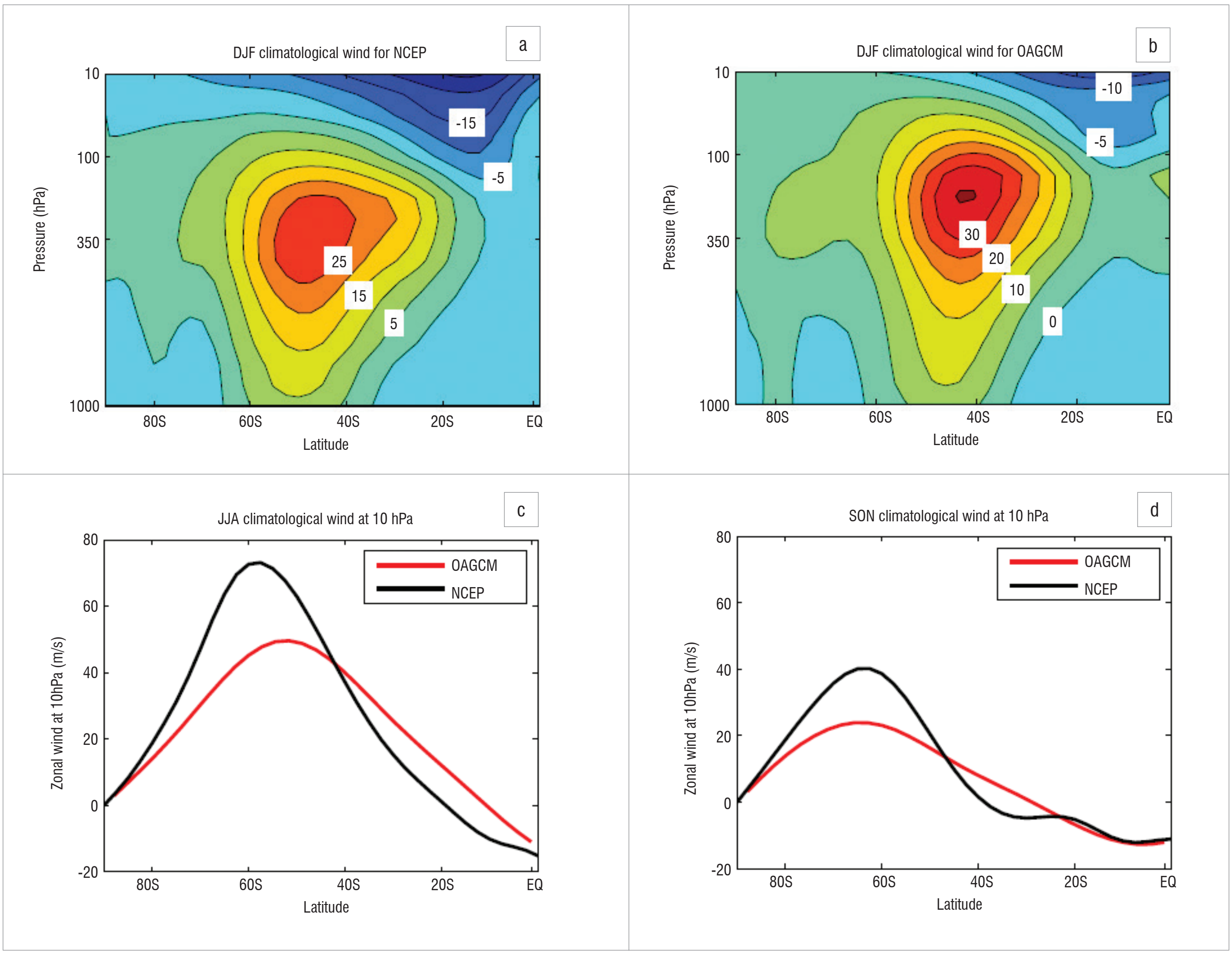


Figure 1: Zonal wind component averaged in the longitudinal direction for (a) NCEP and (b) OAGCM at all isobaric levels during December-January-February (DJF) and at 10 hPa during (c) June-July-August (JJA) and (d) September-October-November (SON). (a and b) Summer (DJF) tropospheric wind response to the strength of stratospheric polar vortex during (c) winter (JJA) and (d) spring (SON). Red contours in (c) and (d) represent OAGCM data and black contours represent NCEP data. Wind speed is measured in m/s.

Evolution of polar cap temperatures

We now consider the climatological evolution of polar cap temperatures. They are defined as zonally averaged (longitude is considered negligible) temperature fields that are also averaged from 70° to 90°S. The selection of this latitude range confirms the south polar cap and defining the polar cap in this way is common practice.[25,45] Figure 2a and 2b show the polar cap temperature for the NCEP and OAGCM, as a function of month and pressure level. The evolution of lower tropospheric polar cap temperatures is similar in NCEP and OAGCM, but both the lower stratosphere and upper troposphere are quite different. The structure of the changes of NCEP temperatures are consistent with those of ozone concentrations as seen in Atmospheric Chemistry and Climate (AC&C) and Stratospheric Processes And their Role in Climate (SPARC) data (Figure 1).[45] Because lower stratospheric ozone has not depleted in the OAGCM, as opposed to in NCEP reanalysis which has a better representation of ozone, the lower stratosphere of the former is much warmer than that of the latter. By the thermal wind relation[34] one can see why the NCEP polar vortex is stronger than its OAGCM counterpart (Figure 1).

These results reveal that the NCEP climatological structure of the zonal wind distribution in the lower stratosphere during winter and the associated summer zonal wind distribution throughout the tropospheric mid-latitude regions are different from those of the OAGCM. The relative strengths of the polar night jet in the respective data sets are also consistent with the climatological temperatures in the lower stratosphere and upper troposphere, which in turn are consistent with ozone representation in the data sets.

Seasonal trends

In the discussion above we demonstrated that there are significant differences between the mean stratospheric and mean tropospheric winds and temperatures during the winter, summer and spring seasons in NCEP and OAGCM data. It was further suggested that this finding might be linked to the inadequate radiative forcing in the OAGCM, which may be caused by unrealistic ozone variations. In this section we consider the changes in the lower stratospheric temperatures during the spring season in the different data sets and their associated tropospheric response during the summer months, as established in previous modelling and observation studies. This work is aimed at demonstrating the likely importance of proper radiative forcing representation in seasonal prediction systems. The polar cap stratosphere is climatologically colder in the NCEP case than in the OAGCM, as noted above.[34]

These climatological thermal structures could be associated with changes in the temperatures. As shown in Figure 3a, NCEP polar cap temperature exhibits cooling in the lower stratosphere and upper troposphere during October to March. This result is consistent with observational studies[18] which used radiosonde data sets and model simulations[19,25] using chemistry climate models with state-of-the-art interactive stratospheric

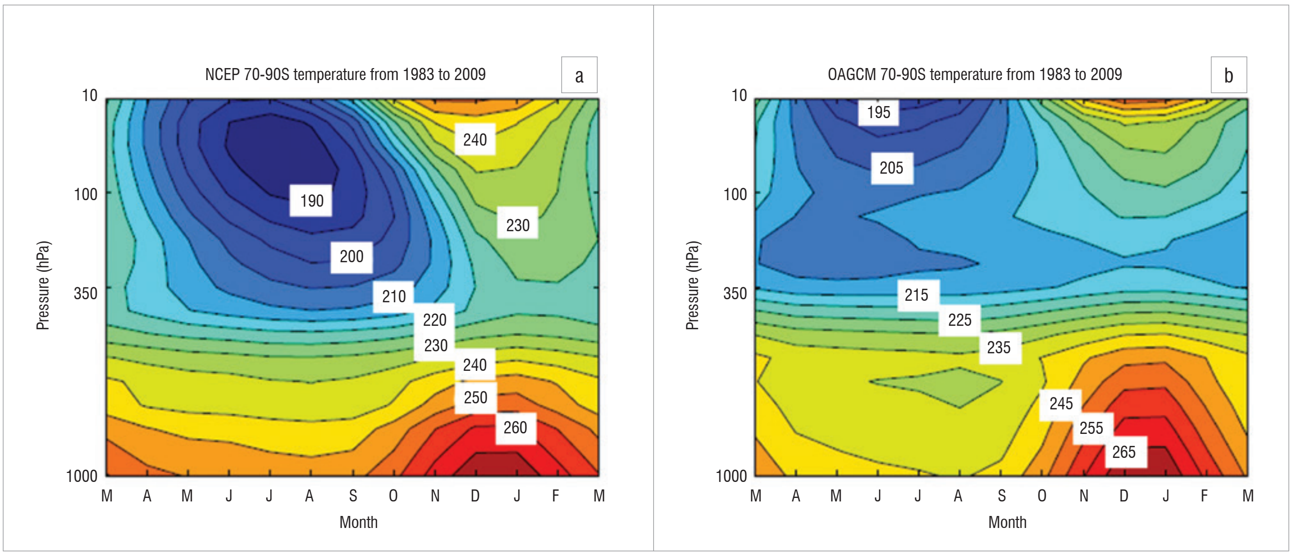


Figure 2: Polar cap temperature evolution as a function of month and pressure for (a) NCEP and (b) OAGCM. The contour intervals are 5 K.

chemistry schemes. As shown in these studies and others, this cooling is a direct result of ozone depletion with increasing GHG concentrations augmenting it.[17] There is no such cooling in the OAGCM polar cap (Figure 3b). The reason for this absence of cooling is that the ozone does not deplete in the model because, as noted previously, it is represented as monthly climatologies. Furthermore, the GHG concentrations do not increase in the OAGCM. Figure 3c shows the corresponding linear trends in NCEP zonal winds averaged between 50˚S and 70˚S as a function of month. The zonal wind response to the cooling begins during October in the lower stratosphere and upper troposphere. This response is shown by positive linear trends and it is most evident in the middle and lower troposphere during the summer months (Figure 3c) – a phenomenon that is well documented. As expected, a similar response does not occur in the OAGCM. Instead, there is a negative trend in the zonal wind between 50°S and 70°S in the OAGCM (Figure 3d). This result suggests (as will be further elaborated on below) that the tropospheric response in the model, although consistent with the associated polar cap temperatures, is not realistic.

Impact of ozone concentration

It has been demonstrated in Figure 3 that when ozone variations are represented in a realistic manner (meaning that ozone depletion actually occurs in the NCEP data sets), then the stratospheric cooling that results from ozone depletion has a tropospheric response that is particularly evident during austral summer. If this is not the case, as shown by the coupled model, ozone does not deplete and thus an opposite tropospheric response occurs. These features are shown as positive and negative linear trends represented by thin black and red contours, respectively (Figure 4a and 4b). This tropospheric response is actually an acceleration (or deceleration) of the eddy-driven jet on the poleward (or equatorward) side and is shown in Figure 4a. This response is consistent with what is seen in Figure 3c, which shows positive linear trends as a result of stratospheric cooling caused by ozone depletion (Figure 3a). A response in December-January-February mainly occurs as a result of stratospheric anomalies that have a time lag of a few months to descend to the surface.[18,46]

The linear trends are a manifestation of the poleward movement of the eddy-driven jet as shown by the red linear trend line in Figure 4c, which moved from about 46°S to about 50°S during the analysed period. As noted before, the shift in the jet is a result of a cooler ozone-induced polar stratosphere (Figure 3a). A Monte Carlo or re-randomisation test[47,48] was performed on the reanalysis data linear trend of Figure 4c to test for its statistical significance by randomly creating a time series from the original interannual NCEP variations. After each re-randomisation, a least squares regression line was fitted to the randomised data from which the trend was calculated, and the process was repeated 10 000 times. The number of times the original trend was larger than the re-randomised trends was noted. Less than 1% of the re-randomised trends were larger than the original trend. This sloping linear fit is therefore statistically significant at the 99% level of confidence.

The magnitude of the poleward shift can also be measured using the climatological position as a reference (green straight line in Figure 4c). The initial sub-climatological jet position was equatorward of the climatological position and ended up on the poleward side of it at the end of the study period. The OAGCM jet tended to decelerate (accelerate) on the poleward (equatorward) side, although this tendency was weak (Figure 4b). The response of the jet in Figure 4b is also consistent with the tropospheric response seen in Figure 3d. These responses are a result of the warm stratosphere caused by the non-depletion of ozone. Moreover, this weak tropospheric response has resulted in the lack of trends shown by the black straight line coinciding with the climatological jet position in Figure 4c. It is clear that there are notable differences between NCEP and OAGCM lower stratospheric cooling during the spring and the associated summer tropospheric response to that cooling.

To conclude, we point to the possible importance of the proper lower stratospheric ozone variations to seasonal prediction by considering the interannual variability of the position of the jet. As noted in the introduction, variations of the jet position relative to South Africa are important to the country's summer rainfall. Most of the summer rainfall results from long cloud bands that stretch diagonally across the mainland and connect the tropical processes to the mid-latitudes through tropical temperate troughs.[49] When the eddy-driven jet is placed anomalously equatorward, these cloud bands become displaced and occur outside the eastern boundaries of South Africa, which leads to a dry summer season.[12]

Summary and conclusions

It has been demonstrated that in an OAGCM that has a climatological representation of lower stratospheric ozone, the depletion thereof (as opposed to in the reanalysis data) does not occur. The non-depletion of ozone leads to a lack of stratospheric temperature cooling, in contrast with the findings of previous studies[18,19,25] that have shown that ozone depletion leads to the cooling of the lower stratosphere. The reanalysis data show a significant austral summer tropospheric response that manifests as an acceleration of the eddy-driven jet on the poleward side and its deceleration on the equatorward side. These features are caused by a gradual and persistent poleward migration of the sub-climatological jet core. Because of the lack of ozone depletion in the OAGCM, the same is not observed. Instead, the sub-climatological jet remains largely stationary as indicated by the weak acceleration and deceleration of the climatological jet stream on the equatorward and

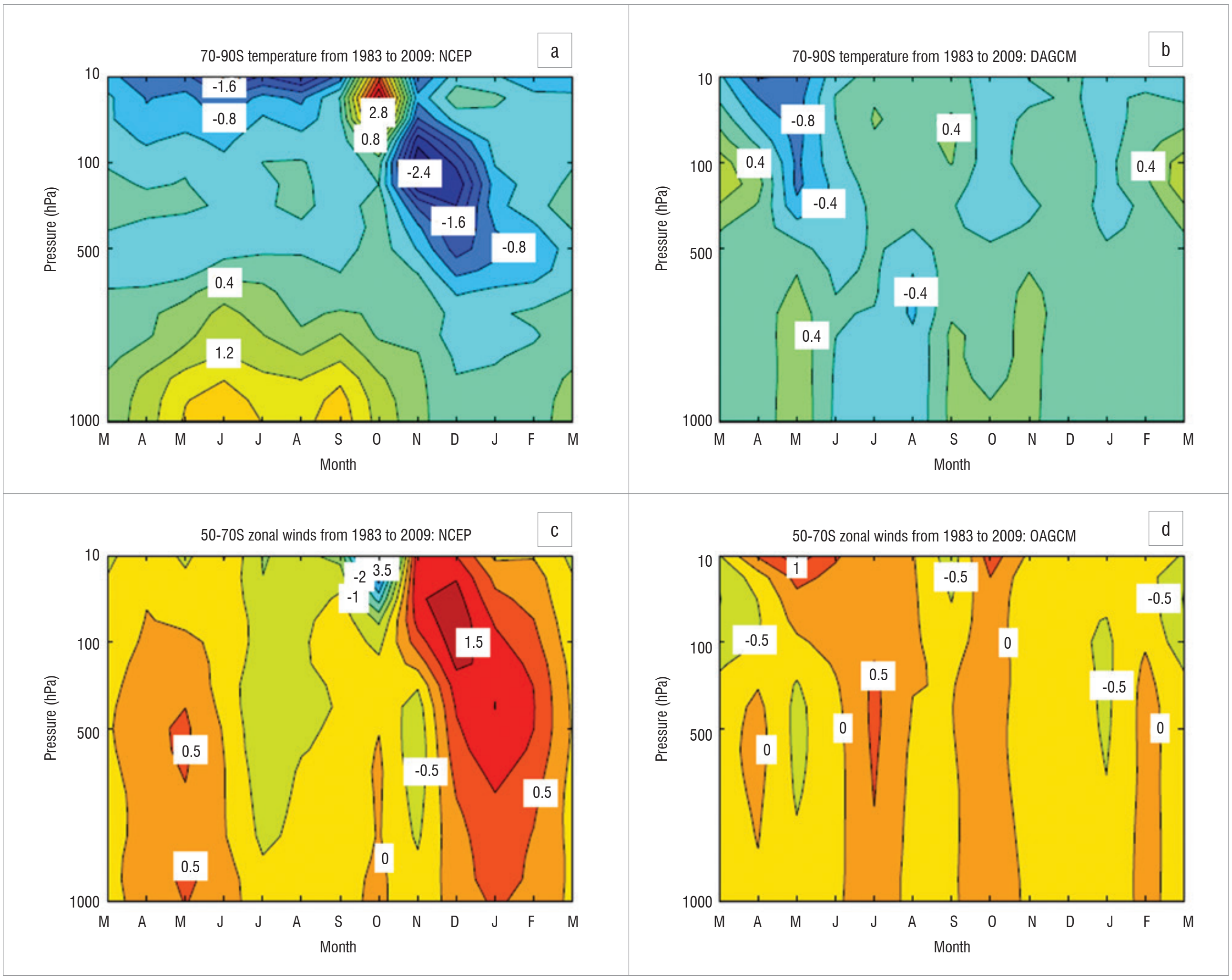


Figure 3: Linear trends in polar cap (averaged over 70–90°S) temperature measured in K/decade from 1983 to 2009 in (a) NCEP and (b) OAGCM data sets. Mid-latitude wind (averaged from 50°S to 70°S) linear trends measured in m/s/decade. The contour intervals in the top and bottom panels are 5 K/decade and 0.2 m/s/decade, respectively.

poleward sides, respectively. However, the aim of this study was to demonstrate that the exclusion of long-term changes in stratospheric ozone (and GHGs) leads to an inaccurate position of the jet stream and the interannual vacillations of the zonal wind fields occur about an inaccurate latitude in the OAGCM – far more north than where it is supposed to be. However, the phases of the zonal wind anomalies observed in the NCEP reanalysis were correctly reproduced by the OAGCM, suggesting that the model simulated the ENSO signal correctly. The issue of stratospheric wind and temperature anomalies as a source at seasonal predictability under realistic ozone prescription and anthropogenic forcing is beyond the scope of this paper and is currently under investigation.

Whilst the trends in the jet position are associated with ozone depletion, and therefore apply to longer-term stratospheric/tropospheric dynamical coupling, they could also be important for seasonal prediction because South African summer rainfall is regulated by the position of the eddy-driven jet and associated storm tracks, relative to the land.[12] If the jet is placed more poleward than usual, then the country experiences a wet summer season. Otherwise the cloud bands which result in most of the summer rainfall are displaced eastward, leading to a dry summer. Therefore, the ozone depletion induced poleward trends of the jet position (as is observed to be the case in the reanalysis data sets) would affect interannual seasonal rainfall occurrence over South Africa and cause the low frequency vacillations of the jet to be progressively poleward. Other mechanisms such as stratospheric-tropospheric exchange could also influence the movement of this jet because its chemical effect in turn influences the lower stratosphere.[50] However, the exact mechanism by which the stratosphere influences the troposphere is unknown and still is under investigation.[14] In fact, the study by Kang et al.[21] found a direct link between changes in summer rainfall in the subtropical belt and the ozone hole, thus attesting to the importance of this forcing. As there is no such poleward trend in the OAGCM used in this study, its simulation of rainfall cannot be expected to be completely realistic. One way of improving this state of affairs is to improve representation of lower stratospheric ozone as well as GHG concentrations in models. The latter is important in this respect as it cools the lower stratosphere, albeit to a much lesser extent than the ozone hole formation. Advanced modelling centres such as the Canadian Climate Modeling Centre[51] and the National Aeronautics and Space Administration[52] employ interactive stratospheric chemistry. Such configurations offer better simulations than climate models that are prescribed with monthly mean zonal mean ozone because they calculate stratospheric ozone interactively.[24,53] Improvements in the chemistry of the coupled climate model could be facilitated through modelling endeavors such as SPARC and the Chemistry Climate Model Validation (CCMVal). As noted in the CMIP5[28] experiment design, a stratospheric ozone data set is available for inclusion in models operated by centres which do have the capability to implement interactive stratospheric chemistry schemes.

A second recommendation has to do with the way in which the stratospheric dynamics are captured in the OAGCM. Stratospheric/tropospheric coupling is a robust dynamical phenomenon and occurs

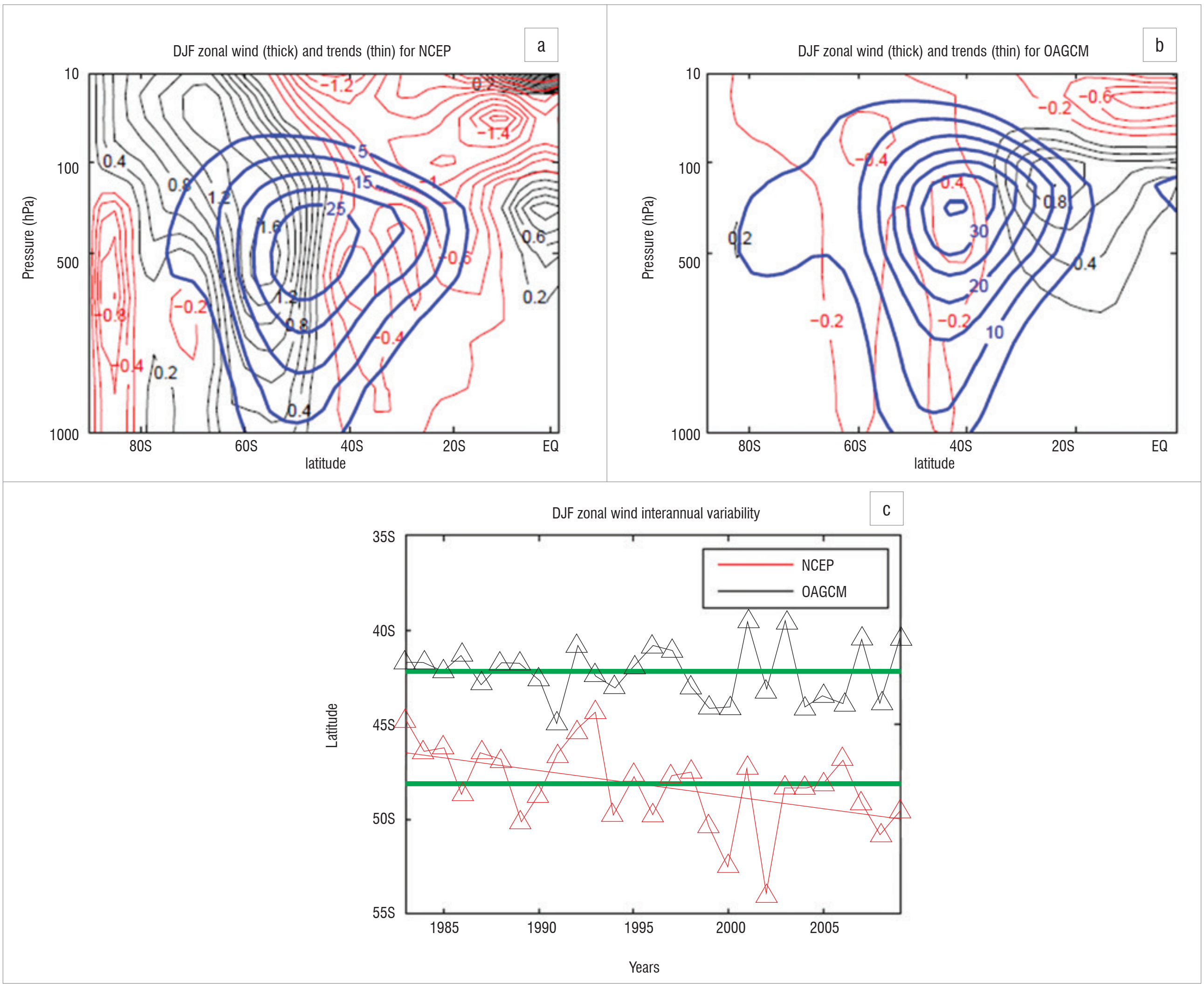


Figure 4: The eddy-driven jet (blue contours) with trends in the zonal winds superimposed for (a) NCEP and (b) OAGCM. Red contours represent negative and black represent positive trends in m/s/decade. The zonal mean isotachs and trends thereof are drawn in 5 m/s and 0.2 m/s/decade contour intervals, respectively. (c) Representation of the interannual variation with NCEP (red line) and OAGCM (black line) linear trends over the analysed period. The green line is the climatological position of the jet in both data sets.

at all timescales.[16] However, it is not inconceivable that if the dynamics of the stratosphere are not properly captured in an OAGCM such as this one studied here, operational seasonal forecasting could be adversely affected. We propose that an OAGCM whose model top is only at 10 hPa would likely be incapable of capturing all the dynamics associated with the variability of the polar vortex. On the basis of this argument, we recommend an increase to the model top to 0.1 hPa so that the stratospheric vortex is captured in its entirety. In addition to this, questions regarding the impact of stratospheric resolution (between 100 hPa and the proposed new model top 0.1 hPa) have not been addressed. It is envisaged that the combination of these efforts could improve seasonal forecasting skill.

The efforts of improving our understanding of the coupled system through modelling and predictability studies should include the knowledge of stratospheric as well as chemical processes (e.g. CO_2 and ozone) which contribute to the so-called 'complete climate system'. This notion was endorsed by the World Climate Research Programme's (WCRP) Climate Variability and Predictability (CLIVAR) in aiming to improve climate and intra-seasonal predictability.[54] The issue of decadal prediction also requires better initialisation of estimates of the current observed atmospheric states in coupled models.[55] However, the advancement of decadal prediction also depends on the improvement of seasonal prediction. [56]

Acknowledgements

We thank Mr Craig Powell, Mr Morne Gijben and two anonymous reviewers for their constructive comments that helped to improve the manuscript. We also thank the Water Research Commission (project number K5/1913) and the Applied Centre for Climate and Earth Systems Science for funding this project.

Authors' contributions

K.M. led the writing of the manuscript. K.M. and T.N. were responsible for the experimental design and analyses and performed the calculations. A.B. made conceptual contributions and prepared the samples (data sets). W.L. calculated the statistical significance of the time series. Finally, T.N., A.B. and W.L. all made contributions that helped improve the original manuscript.

References

1. Goddard L, Mason SJ, Zebiak SE, Ropelewski CF, Basher R, Cane MA. Current approaches to seasonal-to-interannual climate predictions. Int J Climatol. 2001;21:1111–1152. http://dx.doi.org/10.1002/joc.636

2. Landman WA, Goddard L. Statistical recalibration of GCM forecasts over southern Africa using model output statistics. J Climate. 2002;15:2038–2055. http://dx.doi.org/10.1175/1520-0442(2002)015<2038:SROGFO>2.0.CO;2

3. Landman WA, Goddard L. Predicting southern African summer rainfall using a combination of MOS and perfect prognosis. Geophys Res Lett. 2005;32, L15809. http://dx.doi.org/10.1029/2005GL022910

4. Landman WA, Kgatuke MM, Mbedzi M, Beraki A, Bartman A, Du Piesanie A. Performance comparison of some dynamical and empirical downscaling methods for South Africa from a seasonal climate modelling perspective. Int J Climatol. 2009;29:1535–1549. http://dx.doi.org/10.1002/joc.1766

5. Landman WA, Mason SJ, Tyson PD, Tennant WJ. Retroactive skill of multi-tiered forecasts of summer rainfall over southern Africa. Int J Climatol. 2001a;21:1–19. http://dx.doi.org/10.1002/joc.592

6. Mason SJ. Sea-surface temperature-South African rainfall associations 1910–1989. Int J Climatol. 1995;15:119–135. http://dx.doi.org/10.1002/joc.3370150202

7. Landman WA, Mason SJ, Tyson PD, Tennant WJ. Statistical downscaling of GCM simulations to stream flow. J Hydrol. 2001b;252:221–236. http://dx.doi.org/10.1016/S0022-1694(01)00457-7

8. Landman WA, DeWitt D, Lee D, Beraki A, Lotter D. Seasonal rainfall prediction skill over South Africa: One- versus two-tiered forecasting systems. Weather Forecast. 2012;27:489–501. http://dx.doi.org/10.1175/WAF-D-11-00078.1

9. Landman WA, Beraki A. Multi-model forecast skill for mid-summer rainfall over southern Africa. Int J Climatol. 2012;32:303–314. http://dx.doi.org/10.1002/joc.2273

10. Rocha A, Simmonds I. Interannual variability of south-eastern African summer rainfall. Part 1: Relationships with air-sea interaction processes. Int J Climatol. 1997;17:235–265. http://dx.doi.org/10.1002/(SICI)1097-0088(19970315)17:3<235::AID-JOC123>3.0.CO;2-N

11. Hurrell JW, Van Loon H, Shea DJ. The mean state of the troposphere. Meteorology of the southern hemisphere. Meteor Monogr Amer Meteor Soc. 1998;49:1–46.

12. Tyson PD, Preston-Whyte RA. The weather and climate of southern Africa. Cape Town: Oxford University Press; 2000. p. 396.

13. Trenberth KE. Storm tracks in the southern hemisphere. J Atmos Sci. 1991;48:2159–2178. http://dx.doi.org/10.1175/1520-0469(1991)048<2159:STITSH>2.0.CO;2

14. Polvani LM, Kushner PJ. Tropospheric response to stratospheric perturbations in a relatively simple general circulation model. Geophys Res Lett. 2002;29:14284–14287. http://dx.doi.org/10.1029/2001GL014284

15. Kushner PJ, Polvani LM. Stratosphere-troposphere coupling in a relatively simple AGCM: Impact of seasonal cycle. 2006;19:5721–5727.

16. Gerber EP, Buther A, Calvo N, Charlton-Perez A, Giorgetta M, Manzini E, et al. Assessing and understanding the impact of stratospheric dynamics and variability on the Earth system. Bull Amer Meteor Soc. 2012;93:845–859. http://dx.doi.org/10.1175/BAMS-D-11-00145.1

17. Son SW, Gerber EP, Perlwitz J, Polvani LM, Gillett NP, Seo K-H, et al. The impact of stratospheric ozone on southern hemisphere circulation change: A multimodel assessment. J Geophys Res. 2010;115, D00M07. http://dx.doi.org/10.1029/2010JD014271

18. Thompson DWJ, Solomon S. Interpretation of recent southern hemisphere climate change. Science. 2002;296:895–899. http://dx.doi.org/10.1126/science.1069270

19. McLandress CA, Shepherd TG, Scinocca JF, Plummer DA, Sigmond M, Jonsson AI, et al. Separating the dynamical effects of climate change and ozone depletion. Part II: Southern hemisphere troposphere. J Climate. 2011;24:1850–1868. http://dx.doi.org/10.1175/2010JCLI3958.1

20. Marshall GJ. Trends in the southern annular mode from observations and reanalyses. J Climate. 2003;16:4134–4143. http://dx.doi.org/10.1175/1520-0442(2003)016<4134:TITSAM>2.0.CO;2

21. Kang S, Polvani LM, Fyfe JC, Sigmond M. Impact of polar ozone depletion on subtropical precipitation. Science. 2011;332:951–954. http://dx.doi.org/10.1126/science.1202131

22. Feldstein SB. Subtropical rainfall and the Antarctic ozone hole. Science. 2011;332:925–926. http://dx.doi.org/10.1126/science.1206834

23. Perlwitz J. Tug of war on the jet stream. Nat Clim Change. 2011;1:29–31. http://dx.doi.org/10.1038/nclimate1065

24. Perlwitz J, Pawson S, Fogt RL, Nielsen JE, Neff WD. Impact of stratospheric ozone hole recovery on Antarctic climate. Geophys Res Lett. 2008;35, L08714. http://dx.doi.org/10.1029/2008GL033317

25. Waugh DW, Oman L, Newman PA, Stolarski RS, Pawson S, Nielsen JE, et al. Effect of zonal asymmetries in stratospheric ozone on simulations of southern hemisphere climate. Geophys Res Lett. 2009;36, L18701. http://dx.doi.org/10.1029/2009GL040419

26. Purser RJ, Kar SK. Radiative upper-boundary condition or a non-hydrostatic atmosphere. Quart J Roy Meteor Soc. 2002;128:1343–1366. http://dx.doi.org/10.1256/003590002320373328

27. Gerber EP, Polvani LM. Stratosphere-troposphere coupling in a relativey simple AGCM: The importance of stratospheric variability. J Climate. 2008;22:1920–1933. http://dx.doi.org/10.1175/2008JCLI2548.1

28. Taylor KE, Stouffer RJ, Meehl GA. An overview of CMIP5 and the experiment design. Bull Amer Meteor Soc. 2012;93:485–498. http://dx.doi.org/10.1175/BAMS-D-11-00094.1

29. Scaife AA, Spangehl T, Fereday DR, Cubasch U, Langematz U, Akiyoshi H, et al. Climate change projections and stratosphere-troposphere interaction. Clim Dynam. 2012;38:2089–2097. http://dx.doi.org/10.1007/s00382-011-1080-7

30. Roff G, Thompson DWJ, Hendon H. Does increasing model stratospheric resolution improve extended range forecast skill? Geophys Res Lett. 2011;38, L05809. http://dx.doi.org/10.1029/2010GL046515

31. Beraki AF, DeWitt DG, Landman WA, Olivier C. Dynamical seasonal climate prediction using an ocean-atmosphere coupled climate model developed in partnership between South Africa and the IRI. J Climate. 2014;27:1719–1741. http://dx.doi.org/10.1175/JCLI-D-13-00275.1

32. Kanamitsu M, Ebisuzaki W, Woollen J, Yang S-K, Hnilo JJ, Fiorino M, et al. NCEP/DOE AMIP-II reanalysis (R-2). Bull Amer Meteor Soc. 2002;83:1631–1643. http://dx.doi.org/10.1175/BAMS-83-11-1631

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

34. Holton JR. An introduction to dynamic meteorology. 4th ed. Amsterdam: Elsevier Academic Press; 2004.

35. Kim H-K, Lee S. The wave-zonal mean flow interaction in the southern hemisphere. J Atmos Sci. 2004;61:1055–1067. http://dx.doi.org/10.1175/1520-0469(2004)061<1055:TWMFII>2.0.CO;2

36. Postel GA, Hitchman A. A climatology of Rossby wave breaking along the subtropical tropopause. J Atmos Sci. 1999;56:359–373. http://dx.doi.org/10.1175/1520-0469(1999)056<0359:ACORWB>2.0.CO;2

37. Ndarana T, Waugh DW. A climatology of Rossby wave breaking on the southern hemisphere tropopause. J Atmos Sci. 2011;68:798–811. http://dx.doi.org/10.1175/2010JAS3460.1

38. Nakamura H, Shimpo A. Seasonal variations in the southern hemisphere storm tracks and jet streams as revealed in a reanalysis data set. J Climate. 2004;17:1828–1842. http://dx.doi.org/10.1175/1520-0442(2004)017<1828:SVITSH>2.0.CO;2

39. Hitchman MH, Huesmann AS. A seasonal climatology of Rossby wave breaking in the 320-2000-K layer. J Atmos Sci. 2007;64:1922–1940. http://dx.doi.org/10.1175/JAS3927.1

40. Waugh DW, Polvani LM. Stratospheric polar vortices. Geophys Monogr Ser. 2010;190:43–57. http://dx.doi.org/10.2151/jmsj.80.997

41. Waugh DW, Rong PP. Interannual variability in the decay of lower stratospheric Arctic vortices. J Meteorol Soc Jpn. 2002;80:997–1012.

42. Andrews DG, Holton JR, Leovy CB. Middle atmosphere dynamics. Orlando, FL: Academic Press; 1987. p. 489.

43. Nash ER, Newman PA, Rosenfield JE, Schoeberl MR. An objective determination of the polar vortex using Ertel's potential vorticity. J Geophys Res. 1996;101:9471–9478. http://dx.doi.org/10.1029/96JD00066

44. Held IM, Suarez MJ. A proposal for the intercomparison of the dynamical cores of atmospheric general-circulation models. Bull Am Meteorol Soc. 1994;75:1825–1830. http://dx.doi.org/10.1175/1520-0477(1994)075<1825:APFTIO>2.0.CO;2

45. Polvani L, Waugh W, Correa GJ-P, Son SW. Stratospheric ozone depletion: The main driver of 20th century atmospheric circulation changes in the southern hemisphere. J Climate. 2011;24:795–812. http://dx.doi.org/10.1175/2010JCLI3772.1

46. Gillett N, Thompson DWJ. Simulation of recent southern hemisphere climate change. Science. 2003;302:273–275. http://dx.doi.org/10.1126/science.1087440

47. Wilks DS. Statistical methods in the atmospheric sciences. 2nd ed. Amsterdam: Academic Press; 2006. p. 627.

48. Livezey RE, Chen WY. Statistical field significance and its determination by Monte Carlo techniques. Mon Weather Rev. 1983;111:46–59. http://dx.doi.org/10.1175/1520-0493(1983)111<0046:SFSAID>2.0.CO;2

49. Harrison MSJ. A generalized classification of South African summer rain-bearing synoptic systems. J Climatol. 1984;4:547–560. http://dx.doi.org/10.1002/joc.3370040510

50. Holton JR, Haynes PH, McIntyre ME, Douglass AR, Rood RB, Pfister L. Stratosphere-troposphere exchange. Rev Geophys. 1995;33:403–439. http://dx.doi.org/10.1029/95RG02097

51. McLandress CA, Jonsson I, Plummer DA, Reader MC, Scinocca JF, Shepherd TG. Separating the dynamical effects of climate change and ozone depletion. Part I: Southern hemisphere stratosphere. J Climate. 2010;23:5002–5020. http://dx.doi.org/10.1175/2010JCLI3586.1

52. Pawson S, Stolarski RS, Douglass AR, Newman PA, Nielsen JE, Frith SM, et al. Goddard Earth Observing System chemistry-climate model simulations of stratospheric ozone-temperature coupling between 1950 and 2005. J Geophys Res. 2008;113, D12103. http://dx.doi.org/10.1029/2007JD009511

53. Son SW, Polvani LM, Waugh DW, Akiyoshi H, Akiyoshi H, Garcia R, Kinnison D, et al. The impact of stratospheric ozone recovery on the southern hemisphere westerly jet. Science. 2008;320:1486–1489. http://dx.doi.org/10.1126/science.1155939

54. Climate-system Historical Forecast Project (CHFP) [document on the Internet]. No date [cited 2014 Feb 26]. Available from: http://www.wcrp-climate.org/wgsip/chfp/references/CHFP.pdf

55. Hurrel JW, Delworth T, Danabasoglu G, Drange H, Drinkwater K, Griffies S, et al. Decadal climate prediction: Opportunities and challenges. In: Hall J, Harrison DE, Stammer D, editors. Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society; 2009 Sep 21–25; Venice, Italy. ESA Publication WPP-306; 2010. p. 1–23.

56. Goddard L, Hurrell JW, Kirtman BP, Murphy J, Stockdale T, Vera C. Two time scales for the price of one (almost). Bull Amer Meteor Soc. 2012;93:621–629. http://dx.doi.org/10.1175/BAMS-D-11-00220.1

Assessment of atmospheric trace metals in the western Bushveld Igneous Complex, South Africa

AUTHORS:
Pieter G. van Zyl[1]
Johan P. Beukes[1]
Grizelda du Toit[1]
Desmond Mabaso[2,3]
Johan Hendriks[1]
Ville Vakkari[4]
Petri Tiitta[5]
Jacobus J. Pienaar[1]
Markku Kulmala[6]
Lauri Laakso[1,4]

AFFILIATIONS:
[1]Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa
[2]Rustenburg Local Municipality, Rustenburg, South Africa
[3]South African Weather Service, Pretoria, South Africa
[4]Finnish Meteorological Institute, Helsinki, Finland
[5]Department of Environmental Science, University of Eastern Finland, Kuopio, Finland
[6]Department of Physics, University of Helsinki, Helsinki, Finland

CORRESPONDENCE TO:
Pieter van Zyl

EMAIL:
Pieter.vanzyl@nwu.ac.za

POSTAL ADDRESS:
Unit for Environmental Sciences and Management, North-West University, Potchefstroom Campus, Private Bag X6001, Potchefstroom 2520, South Africa

DATES:
Received: 04 Sep. 2013
Revised: 30 Oct. 2013
Accepted: 05 Nov. 2013

KEYWORDS:
air quality; temporal assessment; principal component factor analysis; source apportionment

HOW TO CITE:
Van Zyl PG, Beukes JP, Du Toit G, Mabaso D, Hendriks J, Vakkari V, et al. Assessment of atmospheric trace metals in the western Bushveld Igneous Complex, South Africa. S Afr J Sci. 2014;110(3/4), Art. #2013-0280, 11 pages. http://dx.doi.org/10.1590/sajs.2014/20130280

Trace metal species emitted into the atmosphere from natural and anthropogenic sources can cause various health-related and environmental problems. Limited data exist for atmospheric trace metal concentrations in South Africa, which has the largest industrialised economy in Africa, with significant mining and metallurgical activities. A large fraction of these mineral assets is concentrated in the Bushveld Igneous Complex, with the western limb being the most exploited. To partially address this knowledge gap, atmospheric trace metals were collected in the western Bushveld Igneous Complex at Marikana in the North West Province. Diurnal $PM_{2.5}$ and PM_{10} samples were collected for 1 year. A total of 27 trace metal species were determined. With the exception of Ni, none of the trace metals measured during the sampling period exceeded local or international air quality standard limit values. Total trace metal concentrations in the PM_{10} fraction peaked during the dry months and were regularly washed out during the wet season. A less significant seasonal trend was observed for the trace metal concentrations in the $PM_{2.5}$ fraction; a finding attributed to a faster replenishment of smaller particles into the atmosphere after rain events. About 80% of the PM_{10} trace metal levels measured occurred in the $PM_{2.5}$ fraction, while 40% or more of all metals emanated from the $PM_{2.5}$ fraction. This finding indicated a strong influence of anthropogenic sources. Four meaningful emission sources were determined from explorative principal component factor analysis: crustal, vanadium related, base metal related and ferrochromium related, which correlated well with the anticipated atmospheric trace metal sources in the region.

Introduction

The presence of trace metal species in the atmosphere is mainly attributed to the emission of particulate matter (PM) into the atmosphere by anthropogenic activities and natural sources. Gravimetrically, trace metals represent a relatively small proportion of atmospheric aerosols (generally less than 1%).[1] Natural emissions of trace metals result from different processes acting on crustal minerals, e.g. erosion, surface winds and volcanic eruptions, as well as from natural burning and from the oceans. Sodium (Na), magnesium (Mg), aluminium (Al), potassium (K), calcium (Ca), titanium (Ti), chromium (Cr), iron (Fe) and manganese (Mn) are usually associated with mineral dust and crustal species.[2-5] The predominant anthropogenic sources are pyrometallurgy, biomass burning (veld fires), fossil-fuel combustion and incineration. Atmospheric trace metal species usually associated with anthropogenic activities include arsenic (As), cadmium (Cd), copper (Cu), nickel (Ni), zinc (Zn), vanadium (V), mercury (Hg) and lead (Pb).[6] Metal smelting is regarded as one of the most important anthropogenic trace metal emission sources. During smelting processes, metals in the ores are evaporated from the matrix, which could be emitted into the atmosphere if proper pollution control technology is not applied.[7] Industrial metallurgical processes produce the largest emissions of As, Cd, Cu, Ni and Zn.[8]

Trace metals emitted into the atmosphere can cause a variety of health-related and environmental problems, depending on the extent and time of exposure.[9,10] The potential hazard of several toxic species such as As, Cd, Cr, Hg and Pb is well documented and guidelines for levels of these species in the atmosphere are provided by the World Health Organization (WHO).[11] Trace metals such as Cr, Fe and V have several oxidation states and can therefore participate in many important atmospheric redox reactions[1], which can catalyse the generation of reactive oxygen species that have been associated with direct molecular damage and with the induction of biochemical synthesis pathways[12,13].

Only a few studies on the concentrations of atmospheric trace metals in South Africa have been conducted.[12,14-17] South Africa has the largest industrialised economy in Africa, with significant mining and metallurgical activities. A large fraction of the South African mineral assets is concentrated in the Bushveld Igneous Complex (BIC). In Figure 1, the BIC is indicated on a geographical map of South Africa. The BIC covers an area of 66 000 km^2, which stretches across the North West, Gauteng, Mpumalanga and Limpopo Provinces.[18] This region holds numerous metallurgical operations as a result of large deposits of V, Ni, Cu, Cr, cobalt (Co) and tin (Sn). As much as 80% of the world's platinum group metals (PGMs) are also produced from the BIC.[19-21] Platinum is important in the mitigation of global atmospheric pollutants, because it is mainly used in catalytic converters of vehicular exhaust systems.[17]

The BIC is geographically subdivided into five limbs, of which the western limb is the most exploited and industrially developed.[18] The western BIC also forms part of an air pollution priority area recently declared by the South African Government – the Waterberg-Bojanala National Air Quality Management Priority Area.[22] This declaration indicates the relevance and significance of atmospheric pollution in this region. In this study, atmospheric trace metals were collected in the western BIC at Marikana in the North West Province (Figure 1). Marikana is surrounded by more than 30 mines and 11 pyrometallurgical smelters (indicated in Figure 1) within a 60-km radius, which

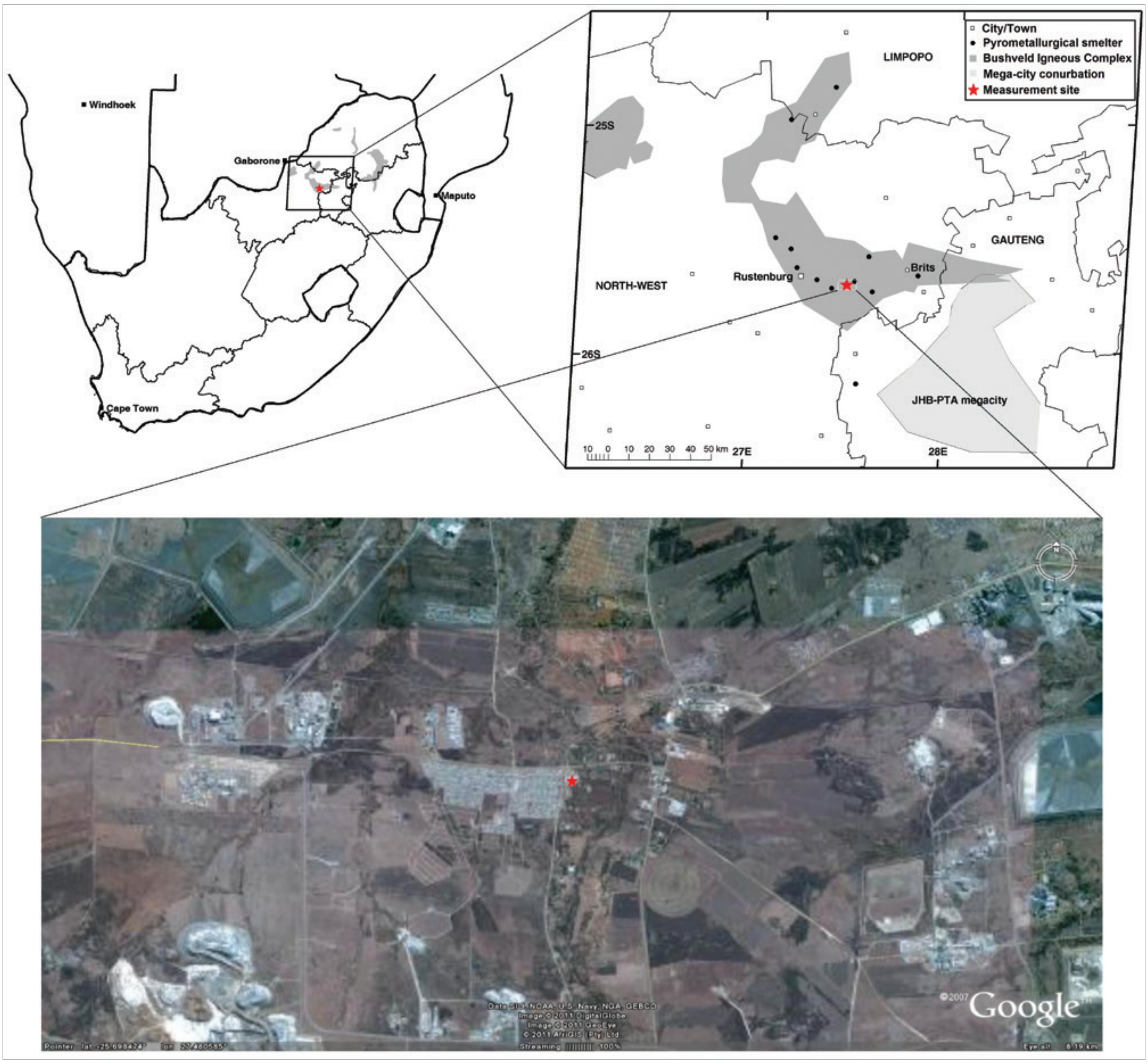


Figure 1: Geographical map of southern Africa and a satellite image indicating the location of the measurement site (25.69845 S, 27.48056 E), as well as the extent of the western Bushveld Igneous Complex and the major point sources.[18] The sampling site is also shown on a satellite image obtained from Google Earth.

include platinum (Pt), Cr, V, Mn and base metal production.[18,23] At present, limited data exist for atmospheric trace metal concentrations in this region.[24] Our general aims in this investigation were to compare trace metal concentrations to existing air quality standards, to conduct a temporal (seasonal and diurnal) and size resolved assessment of trace metal species, as well as to determine possible sources of trace metal species in the western BIC.

Measurement methods and materials

Reagents

Concentrated nitric acid (HNO_3) (CJ CHEM, Johannesburg, South Africa) and concentrated hydrochloric acid (HCl) (Rochelle Chemicals, Johannesburg, South Africa) were utilised for the extraction of metals from samples collected on filters. Metal standards were obtained from Sigma Aldrich (Johannesburg, South Africa) for inductively coupled plasma mass spectrometry (ICP-MS) analysis. All these reagents and standards were analytical grade (AR) chemicals and were used without further purification or processing. Deionised water (18.2 MΩ) was used for all dilutions.

Site description

Measurements were conducted at Marikana (25.69845 S, 27.48056 E, 1170 m AMSL), which is a small village situated approximately 35 km east of Rustenburg, in the North West Province of South Africa (Figure 1). The site was situated in a residential area on the property of the Marikana municipal clinic, which provided access to electrical supply and also ensured the safe keeping of equipment.[18] In Figure 1, the location of the sampling site is also shown on a satellite image obtained from Google Earth. A detailed description of the measurement site and possible sources of pollutant species in the area are presented in Venter et al.[18]

A mobile monitoring station comprehensively equipped with atmospheric measurement instruments was placed at the site. Measurements included meteorological data, radiation measurements, inorganic gaseous species concentrations (SO_2, NO_x, O_3 and CO), as well as physical and chemical properties of aerosol particles. A complete list and detailed description of the atmospheric measurement instruments in the mobile monitoring station used at Marikana have been provided previously.[18,25-27]

Sampling methods

$PM_{2.5}$ and PM_{10} particulate matter samples were collected with MiniVol™ samplers (C&M Consulting Engineers, Pretoria, South Africa) on 2-μm Whatman Teflon® filters (Merck, Johannesburg, South Africa) with a 46.2 mm diameter. The mini-volume samplers were programmed to filter air at a sampling rate of 5 L/min for 12 h/day for a period of 6 days. Each sample contained PM collected for 72 h. The flow rate of the mini-volume samplers was verified once a month with a flow meter. The starting time for sampling was altered every 6 days in order to obtain diurnal samples. Daytime sampling commenced at 06:00, while night-time sampling started at 18:00, corresponding to the approximate light and dark periods, respectively. Sampling was performed for 1 year from November 2008 until October 2009.

The Teflon® filters were carefully inspected prior to sampling to ensure that they were not damaged. Extra care was taken not to damage or contaminate the filters. Filters were handled with latex gloves and tweezers throughout the investigation. The filters were exchanged after each sampling period and stored in airtight Petri dishes in a freezer. Samples were removed from the freezer 24 h before analyses.

Analysis

A surface analysis of sampled filters was performed with an FEI Quanta 2000 scanning electron microscope (SEM) (Hillsboro, OR, USA) with an integrated Oxford Instruments INCA x-sight 400 electron dispersion spectroscope (EDS) (Scotts Valley, CA, USA) microanalysis system.

Atmospheric trace metals were extracted from the collected filters by means of hot acid extraction, as described previously.[28] An ICP-MS analysis of the extracted samples was conducted with an Agilent 7500c series model (Santa Clara, CA, USA). Concentrations of the metals in ambient PM were calculated using the US Environmental Protection Agency (EPA) compendium method IO-3.5.[29] The following 27 trace metal species were detected by means of ICP-MS analysis: Na, Mg, Al, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Cd, Pt, Hg, Pb, beryllium (Be), boron (B), selenium (Se), palladium (Pd), barium (Ba), gold (Au), thallium (Tl), and uranium (U). Metal concentrations that were below the detection limit of the ICP-MS were considered to have concentrations of half the detection limit of the specific metal species considered. This assumption is precautionary and is frequently made in health-related environmental studies.[30]

Statistical analysis

In an attempt to identify possible sources of trace metals measured in this area, principal component factor analysis (PCFA) with Varimax rotation (v. 13.0 SPSS Inc., Chicago, IL, USA) was performed on the data set. PCFA has been used widely in receptor modelling to identify major source categories affecting a given receptor site. The technique operates on sample-to-sample fluctuations of the normalised concentrations. It does not directly yield concentrations of species from various sources, but identifies a minimum number of common factors whose variance often accounts for most of the variance of species.[4,31]

Results and discussion

Trace metal concentrations

The annual average, 3-monthly average and 72-h concentration ranges of all the metals considered are listed in Table 1. Particulate Pd, Hg, Tl and U were below the detection limit of the analytical technique for the entire year. The concentrations of the trace metal species Ca, Co, As, Cd, Ba and Au were above the detection limit for only 25% or less of the time during the sampling period. Concentrations reported for these metals (Pd, Hg, Tl, U Ca, Co, As, Cd, Ba, Au) are therefore likely to be an overestimate because of the precautionary assumption described previously. These metals (shaded in Table 1) are therefore not discussed further as a result of the uncertainties associated with their concentrations.

Atmospheric Fe had the highest annual average concentration of 2.54 μg/m^3, with concentrations ranging between 0.42 μg/m^3 and 19.01 μg/m^3. Mg was the second most abundant species in the atmosphere with an annual average concentration of 2.04 μg/m^3. The atmospheric concentrations of Mg ranged between 0.01 μg/m^3 and 13.6 μg/m^3. Na, B and Al also had relatively high concentrations in the atmosphere with annual average levels of 1.41 μg/m^3, 1.30 μg/m^3 and 1.28 μg/m^3, respectively. The maximum concentration recorded for B was 14.53 μg/m^3, while Na had a maximum level of 7.23 μg/m^3.

In order to regionally and globally compare the trace metal concentrations determined at Marikana, trace metal levels reported in a selection of previous studies in similar environments are also listed in Table 1. Although the aerosol sampling periods and frequencies for most of these previous trace metal studies were not similar to the aerosol sampling period (1 year) and frequency (once a week) in this investigation, these results could be used to contextualise the trace metal concentrations determined at Marikana. Two of these studies were conducted in South Africa – Rustenburg[16] and the Vaal Triangle[12] – while the other investigations were performed in urban and industrial regions in the United Kingdom (UK)[32], Portugal[33] and Argentina[34].

Rustenburg is a city approximately 40 km west of Marikana and is also situated in the western BIC. In one of the studies cited, trace metal samples were collected for a few months at two sites in the Rustenburg region in 2006. In our study, Fe was also the most abundant species with a much higher atmospheric concentration compared to the Fe levels determined at Marikana. Mn and Cr had the second and third highest atmospheric concentrations, respectively. These species were also significantly higher compared to the concentrations of these trace metals measured at Marikana.

The Vaal Triangle is a highly industrialised region approximately 170 km south of Marikana. In the cited study, aerosols were collected in this area at three sites during a 3-day sampling campaign in July 2006 and another 3-day sampling campaign in March 2007. Atmospheric trace metal concentrations measured at the Vaal Triangle were generally lower compared to the trace metal concentrations determined at Marikana. Na had the highest atmospheric concentration and was twofold higher compared to the Marikana measurements of Na. Fe was the second most abundant species in the Vaal Triangle region.

The annual average trace metal concentrations measured in 2009 at the 24 urban and industrial atmospheric monitoring sites of the UK Heavy Metals Monitoring Network are also presented in Table 1. In this monitoring network, samples are obtained for a period of 1 week. A comparison of the atmospheric trace metal concentrations measured at Marikana with these measurements indicates that the trace metal concentrations measured at Marikana were orders of magnitude higher. Trace metal concentrations measured at an industrial region in Lisbon, Portugal in 15 24-h samples collected during November and December 2001 were also much lower. However, the average Ca concentration at this site in Portugal was significantly higher.

The atmospheric trace metal concentrations in this study were also compared to a similar study conducted in an urban and industrial site in La Plata City in Argentina in 1993. Argentina is also considered to be a developing country. Ca concentrations were considerably higher compared to Marikana, where Ca concentrations were below the detection limit of the instrument 25% or less of the time. Similar to Marikana, Fe and Mg were also abundant metal species in La Plata City. Atmospheric concentrations of Mn, Cu, Zn and Pb compared well to the atmospheric levels of these species measured at Marikana.

Comparison to ambient air quality standards

Also indicated in Table 1 are the existing ambient air quality guidelines and standard limit values for trace metal species prescribed by the WHO air quality guidelines for Europe[11], the European Commission Air Quality Standards[35], the National Ambient Air Quality Standards of the US EPA[36] and the National Air Quality Act of the South African Department of Environmental Affairs (DEA).[37] There are currently

Table 1: Annual average, 3-month average and 72-h concentrations of metal species detected. All values are indicated in $\mu g/m^3$. The aerosol trace metal concentrations determined in a selection of previous studies are also presented. Also indicated are the existing ambient air quality standard limit values for trace metal species.

Metal	Annual standard limit	Marikana RSA (2008 – 2009) Annual averages, PM_{10}	Marikana RSA (2008 – 2009) Maximum 3-month averages, PM_{10}	Marikana RSA (2008 – 2009) Concentration range of 72-h samples, PM_{10}	Rustenburg, RSA (2006) PM_{10}[16]	Vaal Triangle, RSA (2006, 2007) PM_{10}[12]	24 Urban and industrial sites in the UK (2009) PM_{10}[32]	Industrial area in Lisbon, Portugal (2001) PM_{10}[33]	Urban/industrial area in La Plata City, Argentina (1993) TSP[34]
Na	-	1.41	2.01	0.13 – 7.23		2.80		1.70	
Mg	-	2.04	2.78	0.09 – 13.57		1.00		0.10	1.47
Al	-	1.28	1.98	0.31 – 2.86				1.90	
K	-	0.68	1.12	0.03 – 3.25		1.30		0.185	
Ca	-	1.08	1.38	0.43 – 4.53				8.80	5.34
Ti	-	0.12	0.28	0.01 – 0.45	0.18	0.02		0.132	
V	1 (b) #	0.04	0.08	0.00 – 0.40	0.16	0.00	0.002	0.043	
Cr	$2.5x10^{4}$ (a) *	0.24	0.68	0.01 – 1.26	1.37	0.05	0.00355		0.00432
Mn	0.15 (a)	0.06	0.10	0.01 – 0.16	4.39	0.12	0.0141	0.03	0.0255
Fe	-	2.54	4.70	0.42 – 19.01	9.76	1.28	0.510	0.35	1.18
Co	-	0.14	0.30	0.01 – 2.42				0.00012	
Ni	0.02 (b)	**0.33**	0.38	0.02 – 1.92	0.77	0.04	0.00256		0.00315
Cu	-	0.18	0.28	0.01 – 0.80	0.21	0.05	0.0153	0.038	0.0295
Zn	-	0.49	0.64	0.04 – 1.48	0.34	0.09	0.0545	0.039	0.273
As	0.006 (b)	**0.26**	0.33	0.00 – 1.08		0.00	0.00069	0.0005	
Cd	0.005 (a),(b)	**0.03**	0.04	0.00 – 0.16		0.00	0.00030		0.00041
Pt	-	0.35	0.46	0.00 – 1.59			0.00		
Hg	1 (a)	0.55	0.70	0.15 – 2.27			0.000486		
Pb	0.5 (a),(b),(d), 0.15 (c)°	0.08	0.14	0.00 – 0.37	0.42	0.04	0.0149		0.0645
Be	-	0.02	0.02	0.00 – 0.07					
B	-	1.30	1.88	0.02 – 14.53					
Se	-	0.58	0.74	0.16 – 2.39				0.0086	
Pd	-	0.41	0.52	0.11 – 1.68					
Ba	-	0.14	0.17	0.00 – 0.73				0.067	
Au	-	0.38	0.48	0.00 – 1.67					
Tl	-	0.27	0.35	0.07 – 1.12					
U	-	0.47	0.60	0.13 – 1.94					

Metal species that were below the detection limit 75% or less of the time during the entire sampling period are shaded; values that exceeded limits of standards are bolded.

[a] WHO air quality guidelines for Europe[11], [b] European Commission Air Quality Standards[35], [c] National Ambient Air Quality Standards of the United States Environmental Protection Agency[36], [d] National Air Quality Act of the South African Department of Environmental Affairs[37]

 # 24-h limit value

** WHO guideline for Cr(VI) concentrations associated with an excess lifetime risk of 1:1 000 000*

° 3-month limit value

TSP, total suspended particulates

guidelines and standards for only seven trace metal species, of which the abovementioned institutions prescribe limit values for only some of these trace metal species. A comparison of the annual average, 3-month average and 72-h concentrations to the appropriate standard limit values indicates that only As and Ni exceeded an existing limit set by the European Commission of Air Quality Standards during the sampling period. The 0.26 μg/m^3 annual average concentration of As and the 0.03 μg/m^3 annual average level of Cd exceeded their annual standard limit values of 0.006 μg/m^{3}and 0.005 μg/m^3, respectively. These results, however, have to be considered within context, because As and Cd were detectable less than 25% of the time during the sampling period, as mentioned previously. The annual average concentration of 0.33 μg/m^3 measured for Ni (range 0.02–1.92 μg/m^3) exceeded the set standard limit value of 0.02 μg/m^3 by an order of magnitude. This finding can possibly be attributed to the metallurgical activities, especially those associated with base metal refining, in the region. This result will be explored further in subsequent sections in this paper. The WHO guideline of 2.5x10^4 μg/m^3 listed for Cr is only for atmospheric concentrations of the Cr(VI) oxidation state with a lifetime risk of 1:1 000 000. The annual average Cr concentration of 0.24 μg/m^3 determined is the total atmospheric Cr concentration, i.e. all the oxidation states of Cr. This result can therefore not be compared to this guideline. V is the only trace metal species with a 24-h standard limit value. Therefore, 72-h samples collected in this study could not be directly compared to this standard limit value. However, the 72-h maximum value of 0.40 μg/m^3 was well below the 24-h limit of 1 μg/m^3.

It is important to note that Pb, which is the only metal species that has a standard limit value prescribed by the South African DEA, did not exceed any of the standards. De-leading of petrol in South Africa could be considered to be partially responsible for these low concentrations. It is also significant to refer to the concentrations of Hg, which were below the detection limit of the analytical instrument for the entire sampling period. This finding was expected, as particulate Hg only forms a small fraction of the total atmospheric Hg, with Hg being mainly present in the atmosphere as gaseous elemental Hg.[38,39] Measurement of ambient Hg concentrations is receiving increasing attention in South Africa and it is likely that a standard limit value for Hg levels will be prescribed in the near future.

Temporal variations

Seasonal trends

Similar seasonal patterns were observed for daytime and night-time measurements. Therefore, the monthly average totals, i.e. daytime and night-time trace metal concentrations, were calculated. The monthly average total trace metal concentrations in the PM_{10} and $PM_{2.5}$ fractions for the 12-month sampling period are shown in Figure 2.

The total trace metal concentrations in the PM_{10} fraction peaked from April to September, with the highest level recorded during June. This observation can at least partially be ascribed to seasonal variations in meteorological parameters. In Figure 3, the precipitation (a) and wind velocity (b) data for the entire sampling period are presented. Although not perfectly correlated, there seems to be a relationship between periods of precipitation and low total atmospheric PM_{10} trace metal concentrations. In the period December 2008 to mid-March 2009, frequent rain events occurred, which resulted in significant atmospheric removal of PM_{10} particles. This relatively wet period was followed by a dry period from the beginning of April 2009 to the middle of October 2009. Rain events again increased from October 2009 up until the end of the sampling period in November 2009.

From Figure 3 it is also evident that the lowest wind speeds were measured from the beginning of May 2009 up until the end of August 2009. This seasonal wind pattern was expected, because it is well known that more stable tropospheric conditions occur during the winter months, while more unstable conditions are associated with summer. Considering strict monovariance conditions, i.e. only wind variation, it can be expected that higher wind velocities would lead to higher PM_{10} trace metal concentrations, especially of dust from crustal species.[2,4] However, in this study, higher wind velocities coincided with wet periods. During these periods, the soil is also wet and dust is not that easily swept from the surface of the earth by strong winds to become airborne. The presented precipitation and wind velocity data indicate that wet removal processes of atmospheric PM_{10} trace metals are more significant than wind generation thereof.

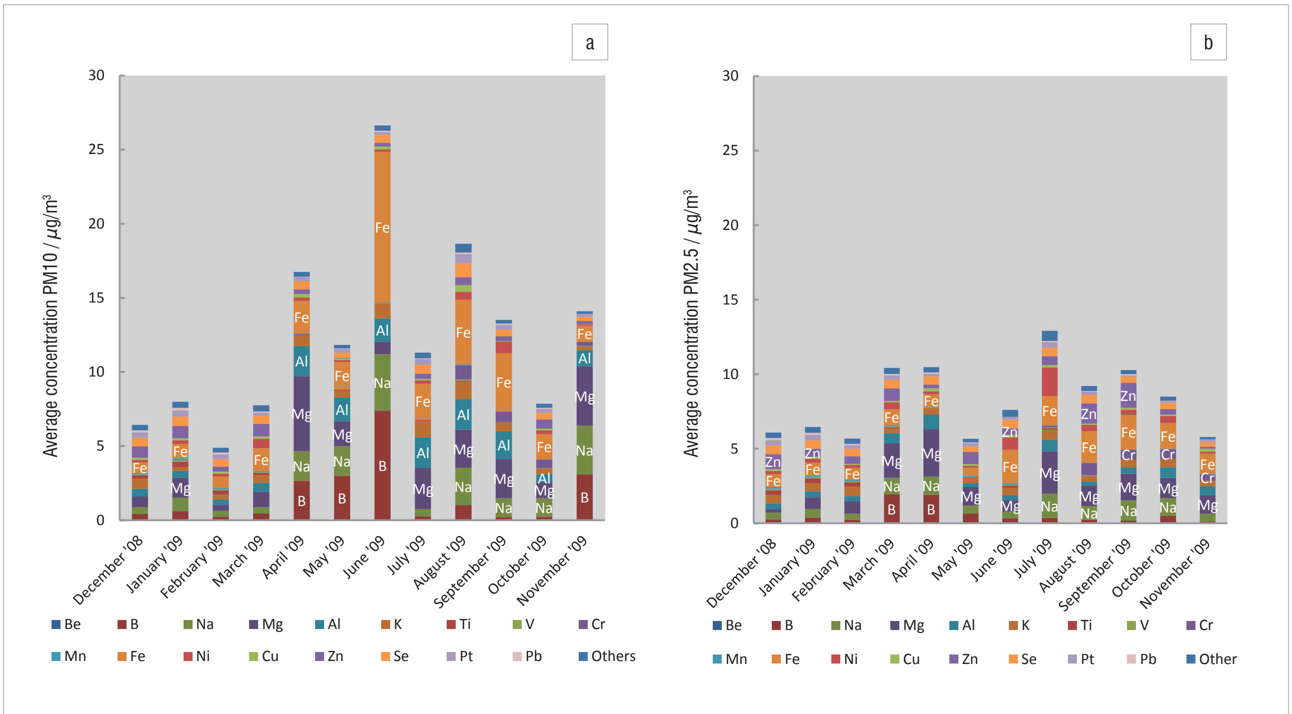


Figure 2: Monthly total trace metal concentration in the (a) PM_{10} and (b) $PM_{2.5}$ fractions.

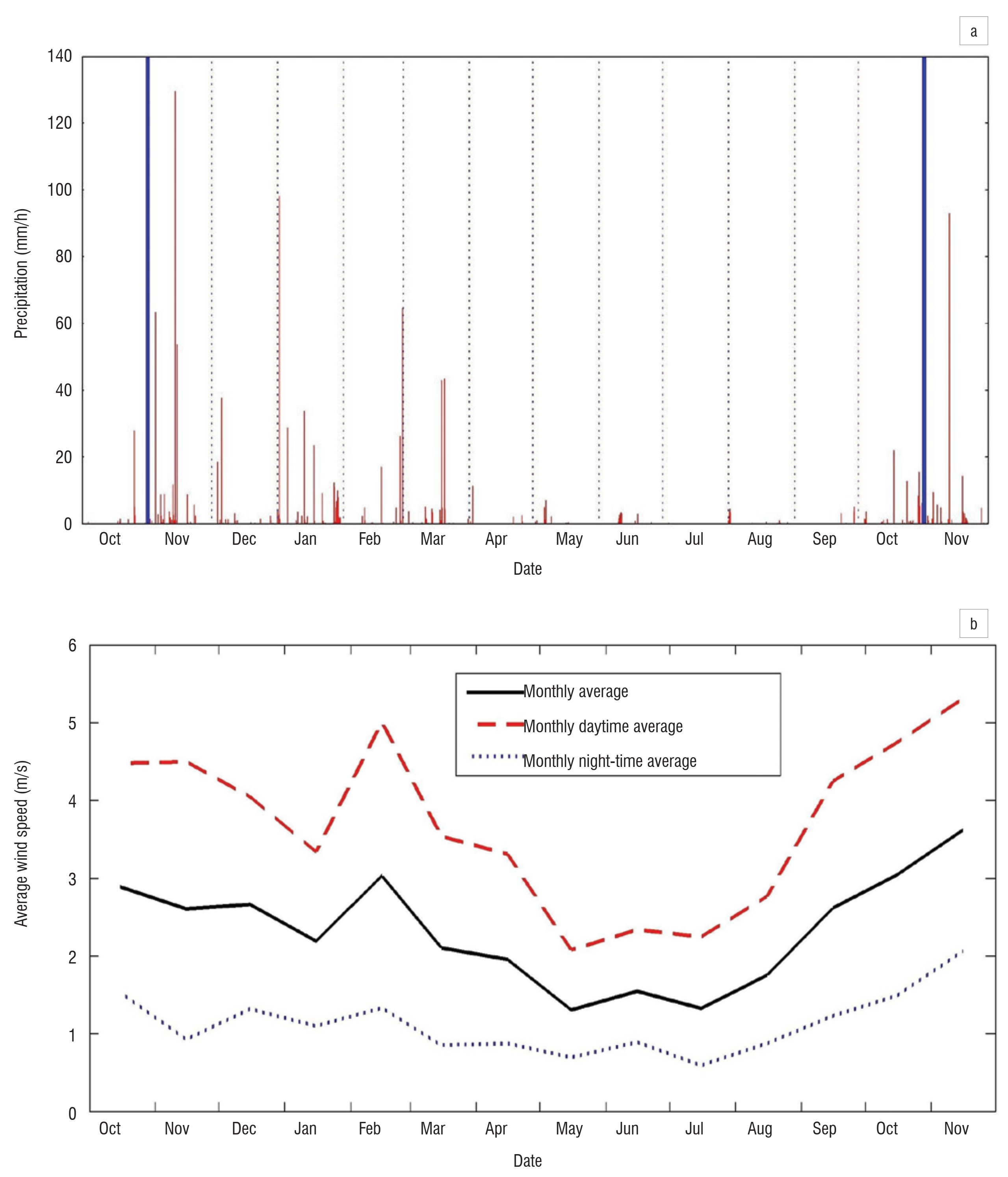


Figure 3: (a) Precipitation events and (b) monthly average wind speeds measured for the entire year of sampling.

A less significant seasonal trend is observed for the trace metal concentrations in the $PM_{2.5}$ fraction (Figure 2). The highest concentrations were determined during the winter month of July. Compared to the chemical composition of atmospheric PM_{10}, a larger fraction of $PM_{2.5}$ usually consists of species associated with chemical processes (natural or anthropogenic). PM_{10} is, for instance, usually strongly associated with windblown dust of crustal species. This implies that dust suppression in the wetter months has a significant influence on PM_{10} concentrations in the atmosphere. Fractionally, a larger portion of the $PM_{2.5}$ composition will originate from processes such as local biomass and domestic burning (household combustion for space heating and cooking), fossil fuel combustion, pyrometallurgical processes and secondary particle formation. Because of the nature of $PM_{2.5}$ sources, smaller particle concentrations are replenished faster than larger particles after washout events. In addition, because measurements were conducted in close proximity of the source region, it is expected that the atmospheric $PM_{2.5}$ concentrations would increase at a faster rate after rain events. Smaller particles also have a longer lifetime in the atmosphere than larger aerosols do.

In Figure 4, monthly normalised trace metal compositions with a breakdown of trace metal species are presented for PM_{10} and $PM_{2.5}$, respectively. From

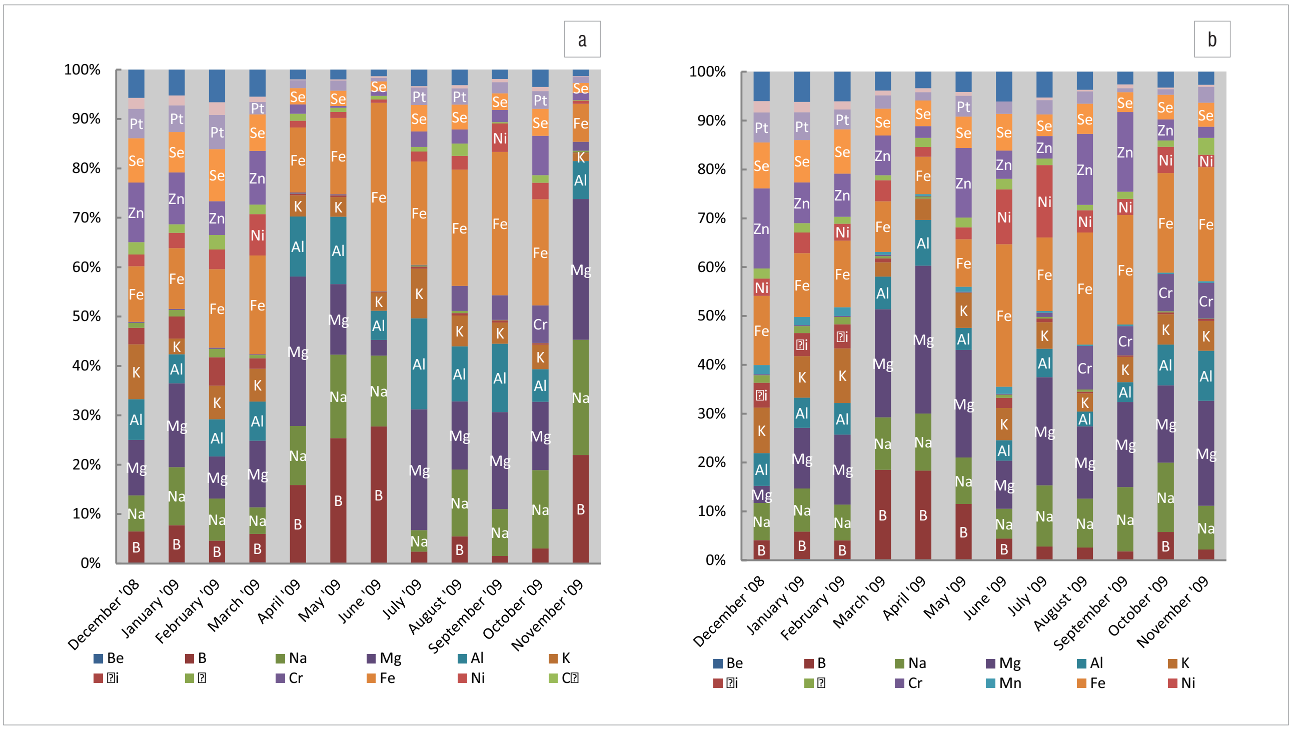


Figure 4: Normalised trace metal distribution in the (a) PM_{10} and (b) $PM_{2.5}$ size fractions.

these figures, no obvious seasonal variations in the fractional trace metal compositions in both particle size fractions are observed. Species that were dominant in both fractions were Fe, Mg, Al, Na and B. It also seems that Zn and Ni are more prevalent in the $PM_{2.5}$ fraction. This size resolved distribution, together with the possible sources of these species, will be discussed later.

Diurnal variations

In Figure 5, the total trace metal concentrations determined for day and night are separated into the $PM_{10\text{-}2.5}$ and $PM_{2.5}$ fractions. $PM_{10\text{-}2.5}$ values were calculated by subtracting the $PM_{2.5}$ total trace metal concentrations from the PM_{10} total trace metal concentrations. From these figures it is evident that $PM_{2.5}$ is more persistent during night-time. During night-time, only 4% of the total trace metal concentrations were in the $PM_{10\text{-}2.5}$ fraction, while 38% of the daytime trace metal concentration was in the $PM_{10\text{-}2.5}$ fraction. In a study conducted at a Nordic background station, Virolahti, the trace metal concentrations of most of the species were also significantly higher in the $PM_{2.5}$ fraction.[40]

Considering that the atmospheric PM_{10} load is expected to be mainly generated by crustal windblown dust, the higher daytime $PM_{10\text{-}2.5}$ trace metal concentration can be explained by significant higher daytime wind speeds compared to night-time throughout the 1-year sampling period as indicated in Figure 3b. Increased traffic volumes during daytime could also possibly lead to higher dust emissions from roads. The higher night-time $PM_{2.5}$ trace metal concentrations can possibly be explained by the formation of low-level inversion layers trapping smaller atmospheric particles. The Highveld of South Africa is well known for its cold and dry winters, with the associated formation of strong low-level inversion layers during the night-time and early morning. These inversion layers usually dissipate in the mid-morning, resulting in the dispersal of PM. The effect (trapping of PM by the inversion layer) is expected to be stronger on smaller particles, as they are more persistent (longer lifetimes) in the atmosphere than larger particles are. Additionally, household combustion for space heating and cooking occurs more during night-time and early mornings, especially during winter when the low-level inversion layers are still prevalent. In a study conducted at three sites in an informal settlement in South Africa, residential coal combustion accounted for 62.1% of the $PM_{2.5}$ measured.[14] Household combustion is usually associated with increased atmospheric $PM_{2.5}$ concentrations.[41]

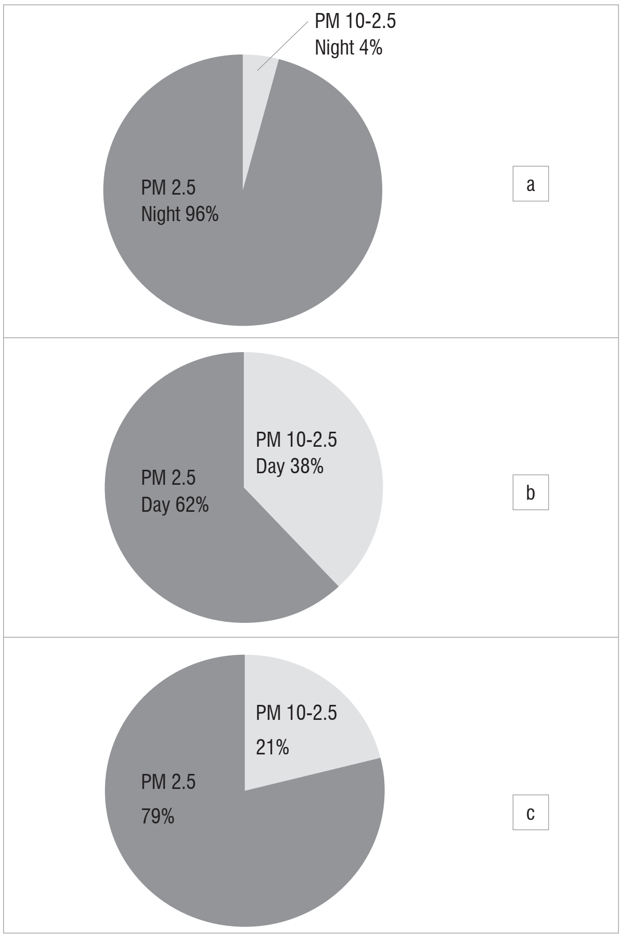


Figure 5: Total trace metal concentrations for (a) day and (b) night separated into the $PM_{10\text{-}2.5}$ and $PM_{2.5}$ fractions. (c) The size distribution of total (day and night) trace metal species separated into the $PM_{10\text{-}2.5}$ and $PM_{2.5}$ fractions.

An air quality assessment of atmospheric measurements conducted between February 2008 and May 2010 at the Marikana sampling site indicated that household combustion was the major source of PM_{10} in this region.[18] These authors[18] also indicated that PM_{10} concentrations peaked in the mornings between 06:00 and 10:00, as well as in the evenings between 17:00 and 22:00. These peak periods could potentially have an influence on atmospheric trace metal concentrations. However, the influence of these peak concentrations on trace metal concentrations measured in this study could not be quantified, as these peaks overlapped with the daytime and night-time sampling periods during which PM was collected for trace metal measurements.

Also presented in Figure 5 are the total trace metal contributions (day and night combined) of the $PM_{10\text{-}2.5}$ and $PM_{2.5}$ fractions. A total 79% of the total trace metal concentrations that were determined were present in the $PM_{2.5}$ fraction. This prevalence of trace metals in the $PM_{2.5}$ section is indicative of the existence of significant anthropogenic sources in addition to natural processes.

Size distribution of trace metals

Each trace metal species (detected above the detection limit) was also fractionalised separately between $PM_{10\text{-}2.5}$ and $PM_{2.5}$. These distributions are shown in Figure 6. From these results it is evident that for all the metal species that occurred at levels above the detection limit, 40% and more of the specific metal were present in the $PM_{2.5}$ fraction. In contrast, the $PM_{10\text{-}2.5}$ fraction comprised only Al, B, Fe, Na, K, Pt, Cu, Mg, Be, Pb, Se and Ti. As mentioned previously, the $PM_{10\text{-}2.5}$ fraction is usually associated with windblown dust, which is usually associated with Al, Fe, Na, Mg and Ti.[2-5] Measurements conducted at a Nordic background station in Virolahti showed significantly smaller fractions of Al, Fe and Mg occurring in the $PM_{2.5}$ size ranges compared to measurements of these species at Marikana.[40] This also signifies the strong influence of industrial activities on atmospheric trace metal concentrations in the western BIC. Cr, Mn, V, Zn and Ni, which are typically associated with metallurgical activities,[6,19-21,42] occurred almost exclusively in the $PM_{2.5}$ fraction. The absence of these species in the coarse fraction implies the improbability of these species originating from windblown dust at this specific site. It is significant to note that pyrometallurgical smelters related directly to the production of Cr, V and Ni are common anthropogenic activities in the western BIC. It is also well known that Zn and Mn are volatilised during pyrometallurgical smelting processes.[42] The percentages of these trace metals occurring in the $PM_{10\text{-}2.5}$ fraction are relatively high compared to studies conducted in urban areas in Central England[43], as well as in the northern and northwestern parts of Switzerland.[44]

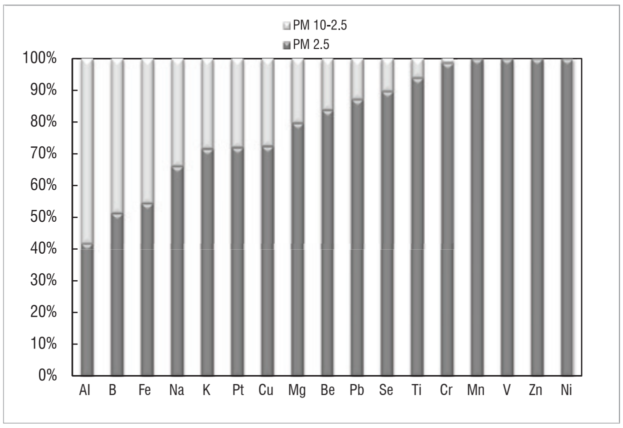


Figure 6: Size distributions of the individual metal species detected.

SEM-EDS analysis

Backscatter micrographs of a typical PM_{10} sampled filter are shown in Figure 7a and 7b. Particles collected are clearly visible in these micrographs. The shapes of particles can be used as an indicative method or observation to distinguish between windblown particles and particles originating from pyrometallurgical processes. Beukes at al.[21,45] reported that small atmospheric rounded particles are likely to originate from the off-gas of pyrometallurgical processes. In Figure 7b, typical rounded and unevenly shaped particles are indicated with white circles and white squares, respectively. This visual observation emphasises the deductions made from chemical analyses described in previous paragraphs that particles are likely to originate from anthropogenic activities (e.g. pyrometallurgical processes), as well as from windblown dust.

In addition to visual inspection, SEM-EDS analysis was used to obtain the surface chemical composition. SEM-EDS is a semi-quantitative technique. This technique confirmed the presence of most of the trace metal species identified with ICP-MS analysis. Additionally, non-metallic species were also detected, of which S was the most significant. The pyrometallurgical industry associated with the large PGM industry in the western BIC is well known for high SO_2 emissions. In general, the S content was more than 6% of the total elemental weight percentage, excluding C and F, which are the elements associated with the Teflon filter media. This relatively high S content clearly indicates the strong influence of pyrometallurgical activities on the PM sampled. This finding is another confirmation of conclusions made previously with regard to the influence of anthropogenic activities on the trace metal species measured.

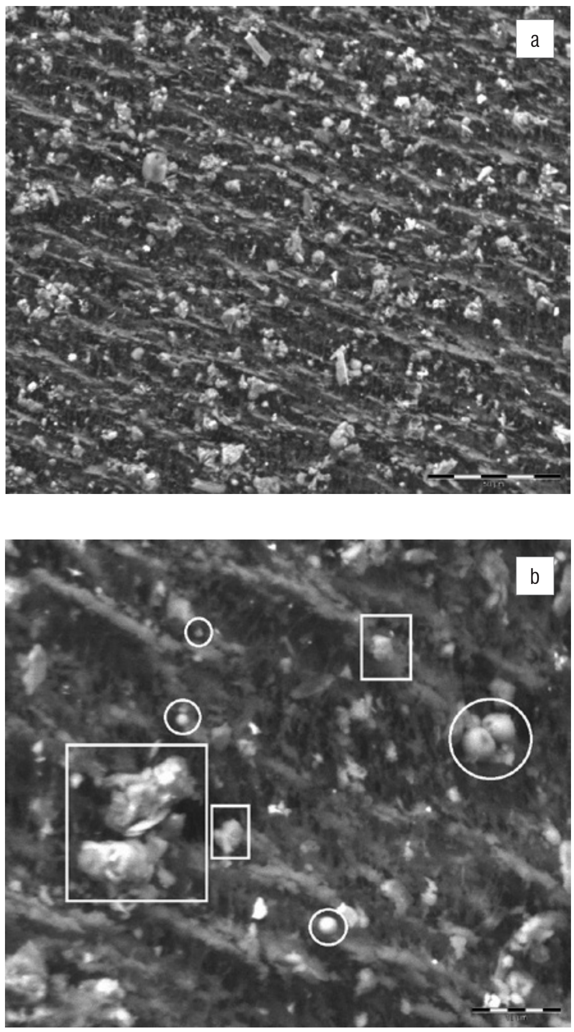


Figure 7: Backscatter micrographs of a typical PM_{10} filter. (a) The SEM enlargement from which electron dispersion spectroscopy analysis was performed and (b) a zoomed-in micrograph of the same surface indicating shapes of specific particulates.

Source apportionment (factor analysis)

Principal component factor analysis was applied as an explorative tool in this investigation. Although ours is one of the most comprehensive data sets for atmospheric trace metals in South Africa, from a statistical perspective the data set was not large enough for definitive source apportionment with PCFA. Factor analysis of smaller data sets is more susceptible to statistical artefacts, such as autocorrelation or the influence of outliers. This implies that some correlations are not necessarily considerable. Therefore, only the most apparent groupings of metal species relating to expected sources in the region were identified.

The data set used for PCFA was the combined ambient concentrations determined for all 27 metal species in both the $PM_{2.5}$ and PM_{10} fractions. The data set was subjected to Box-Cox transformation in an effort to eliminate some possible artefacts. There are no well-defined rules on the number of factors to be retained. Generally, factors that are either meaningful or with eigenvalues larger than one are kept.[4,5,46] Factor loadings obtained for the data sets are presented in Figure 8. Six factors were retained from the transformed data set that explained 78.9% of the variance.

Factor 1 obtained is mostly loaded with the metal species Pd, Hg, Tl, U Ca, Co, As, Cd, Ba and Au, which were species that were above the detection limit 25% or less of the time. Therefore, this factor was excluded in the elucidation of possible sources of trace metal species. Factor 5 obtained for this data set has high loadings of B and Na, as well as moderately high loadings of Al and Mg. These trace metals might be indicators of crustal sources, which are more likely to be correlated in Factor 4. Therefore, Factor 5 was considered as an unidentified source. Factor 1 and Factor 5 explained 34.5% and 6.1% of the total system variance, respectively. Four meaningful sources were identified from the factor analysis of the transformed data set and are discussed below.

Factor 2 of the modified data set was highly loaded by Ti, V and Mn. This factor explains 13.9% of the variance and was identified as vanadium-related sources. There is at least one large V smelter in the western BIC; therefore, the presence of V in the atmosphere in this region can be ascribed to pyrometallurgical processes. A correlation that was expected for atmospheric V was its association with Ti, because titaniferous magnetite deposits are used for V recovery and have relatively high Ti content.[47] Mn is present in most of the ores from which metals (Cr, V, etc.) are produced in the western BIC. Mn has a substantially lower vapour pressure than most of the heavy metals produced in this region and is therefore more volatile.[42] Many of the pyrometallurgical industries will therefore be sources of atmospheric Mn. Although it is difficult to provide an exact explanation for the correlation of Mn with V and Ti as a result of the complexity of sources in this region, a correspondence to pyrometallurgical sources is indicated.

Factor 3 of the data set correlated to Mn, Fe, Co, Ni and Zn, which is indicative of base metals produced in this region. These base metal smelters refine base metals that were extracted during PGM processes. This factor explains 10.2% of the variances of the data. This factor was identified as the base metal industry-related sources.

Factor 4 explains 10.0% of the total system variance and was loaded with Mg, Al, K, Fe, Cu and Pt. Explorative investigation revealed that this factor is mainly loaded with metal species that are considered to be from windblown dust, i.e. Mg, Al, K and Fe. Therefore, this factor was identified as the crustal factor.

Another factor identified was ferrochromium-related pyrometallurgical sources. This factor was calculated in Factor 6 of the data set and has high loadings of Cr and Fe. It also correlated to Ni, Pt, Na and K. Factor 6 accounts for 4.2% of the variances in the data set. A correlation between Fe and Cr was anticipated because of the presence of ferrochromium smelters in the region, as described in Venter et al.[18]

Conclusions

Atmospheric Fe had the highest annual average concentration, while Mg was the second most abundant species in the atmosphere. Relatively higher levels were also observed for Na, B and Al. With the exception of Ni, none of the trace metals measured at Marikana during the sampling period exceeded local and international standard limit values. The relatively high Ni concentration could possibly be attributed to base metal refining in the region. Pb, which is the only metal species that has a standard prescribed by the South African DEA, did not exceed any of the standard limit values. It is also significant to mention that Hg was below the detection limit of the analytical instrument for the entire sampling period; a finding that was expected because particulate Hg only forms a small part of the total atmospheric Hg.

The wet removal processes of atmospheric PM_{10} trace metals were more significant than wind generation thereof. Total trace metal concentrations

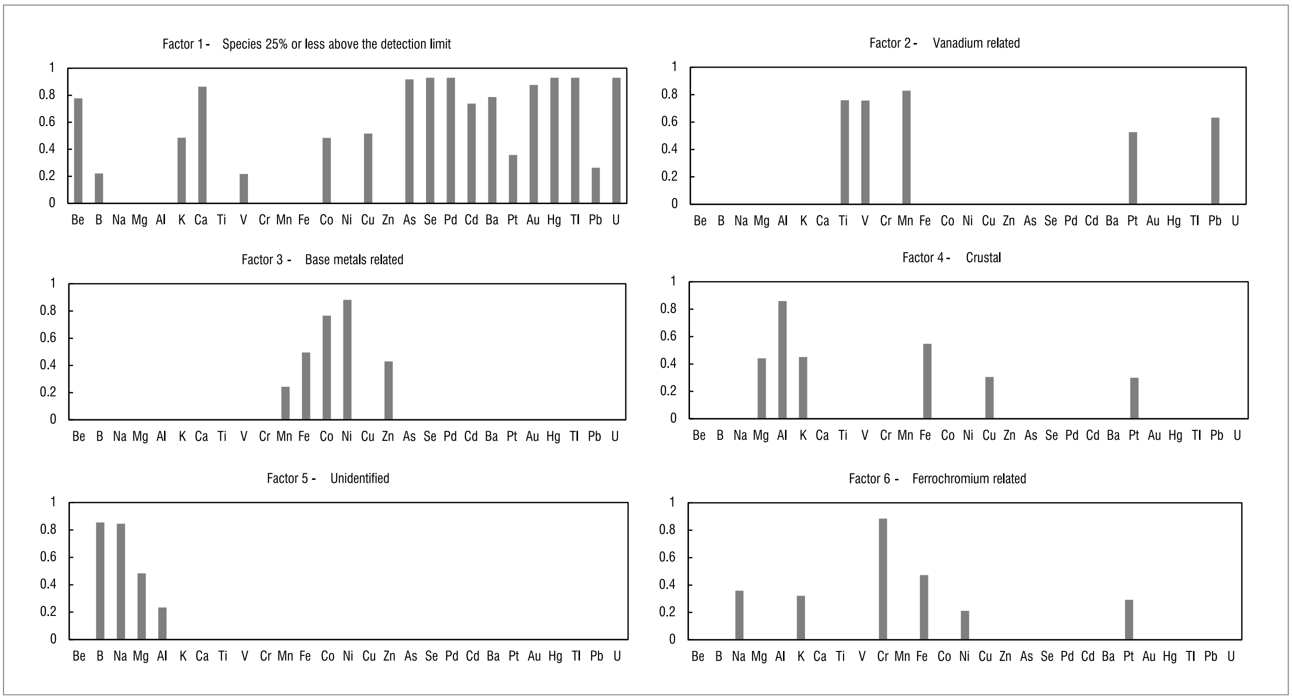


Figure 8: Principal component factor analysis with Varimax rotation of Box-Cox transformed data.

in the PM_{10} fraction peaked during the dry months and were mostly washed out during the wet season. A less significant seasonal trend was observed for the trace metal concentrations in the $PM_{2.5}$ fraction, which was attributed to a faster replenishment of smaller particles into the atmosphere after rain events, as well as the close proximity of the sampling site to the major $PM_{2.5}$ sources. This finding indicates the anthropogenic origin of the smaller particles.

Separation of trace metal concentrations into $PM_{10-2.5}$ and $PM_{2.5}$ fractions indicated that, during night-time, only 4% of the total trace metal concentration was in the $PM_{10-2.5}$ fraction, while the $PM_{10-2.5}$ fraction contributed to 38% of the total daytime trace metal concentrations. An air quality assessment for atmospheric measurements conducted between February 2008 and May 2010 at the Marikana sampling site indicated that PM_{10} concentrations peaked in the early mornings and evenings. These peak periods could potentially have an influence on atmospheric trace metal concentrations. However, the impact of these peak periods on trace metal concentrations measured in this study could not be quantified, as the daytime and night-time sampling periods during which PM was collected for trace metal measurements overlap with these peak periods.

A total of 79% of the total trace metal levels that were measured were in the $PM_{2.5}$ fraction, which indicates a strong influence of anthropogenic sources. Fractionalisation of each of the trace metal species detected showed that for each metal species, 40% and more of a specific metal was in the $PM_{2.5}$ fraction, with Cr, V, Ni, Zn and Mn occurring almost completely in the $PM_{2.5}$ fraction.

Surface analysis with SEM emphasised results from chemical analysis, which indicated that a large fraction of the particles was likely to originate from anthropogenic activities and from windblown dust. SEM-EDS also detected non-metallic S that is usually associated with the PGM pyrometallurgical industry, which is large in the western BIC.

Explorative factor analysis of the data set for all 27 metals detected resolved four meaningful emission sources – vanadium related, base metal related, crustal and ferrochromium related. These factors correlated very well with known pyrometallurgical sources occurring in the western BIC.

As far as we know, this paper presents the most comprehensive atmospheric trace metal concentration study conducted in South Africa published in the peer-reviewed public domain. Many trace metal studies performed in South Africa have been carried out by local authorities or certain industries and are not available in the public domain.[24] Atmospheric trace metal publications available in peer-reviewed literature are limited to shorter measurement periods and reduced frequencies of sampling with fewer trace metal species identified[12,14,16], or only report on specific trace metals such as Pb and Hg[15,17].

Acknowledgements

We acknowledge the Marikana municipal clinic for providing the electricity supply and security for the mobile monitoring station. We also thank Dr Louwrens Tiedt for assistance with SEM-EDS analysis.

Authors' contributions

P.G.v.Z., J.P.B. and G.d.T. were the main investigators in this study. P.G.v.Z. and J.P.B. were the project leaders and wrote the manuscript. G.d.T. conducted this study as part of her MSc degree and performed most of the experimental work. P.G.v.Z. and J.P.B. were also study leaders of the MSc. D.M. assisted logistically in collecting the particulate matter samples at Marikana. J.H. assisted in performing the ICP-MS analysis of the collected samples. V.V., P.T., J.J.P., M.K. and L.L. made conceptual contributions.

References

1. Colbeck I. Environmental chemistry of aerosols. Oxford: Blackwell; 2008. http://dx.doi.org/10.1002/9781444305388
2. Rastogi N, Sarin MM Quantitative chemical composition and characteristics of aerosols over western India: One-year record of temporal variability. Atmos Environ. 2009;43:3481–3488. http://dx.doi.org/10.1016/j.atmosenv.2009.04.030
3. Elefthereriadis K, Colbeck I. Coarse atmospheric aerosol: Size distributions of trace elements. Atmos Environ. 2001;35:5321–5330. http://dx.doi.org/10.1016/S1352-2310(01)00304-1
4. Al-Momani IF, Daradkeh AS, Haj-Hussein AT. Trace elements in daily collected aerosols in Al-Hashimya, central Jordan. Atmos Res. 2005;73:87–100. http://dx.doi.org/10.1016/j.atmosres.2003.09.009
5. Kulkarni P, Chellam S, Flanagan JB, Jayanty RKM. Microwave digestion—ICP-MS for elemental analysis in ambient airborne fine particulate matter: Rare earth elements and validation. Anal Chim Acta. 2007;599:170–176. http://dx.doi.org/10.1016/j.aca.2007.08.014
6. Pacyna JM, Pacyna EG. An assessment of global and regional emissions of trace metals to the atmosphere from anthropogenic sources worldwide. Environ Rev. 2001;9:269–298. http://dx.doi.org/10.1139/a01-012
7. Zheng N, Liu J, Wang Q, Liang Z. Health risk assessment of heavy metal exposure to street dust in the zinc smelting district, Northeast of China. Sci Total Environ. 2010;408:726–733. http://dx.doi.org/10.1016/j.scitotenv.2009.10.075
8. Vassilakos CH, Veros, D, Michopoulus J, Maggosa TH, O'Connor CM. Estimation of selected heavy metals and arsenic in PM10 aerosols in the ambient air of the Greater Athens Area, Greece. J Hazard Mater. 2007;140:389–398. http://dx.doi.org/10.1016/j.jhazmat.2006.11.002
9. Brown LM, Collings N, Harrison RM, Maynard AD, Maynard RL. Ultrafine particles in the atmosphere. London: Imperial College Press; 2003.
10. Karthikeyan S, Balasubramanian R, Iouri K. Particulate air pollution from bushfires: Human exposure and possible health effects. J Toxicol Environ Health A. 2006;69(21):1895–1908. http://dx.doi.org/10.1080/15287390600751264
11. Air quality guidelines for Europe. 2nd ed. World Health Organization Regional Publications, European Series No. 91. Copenhagen: World Health Organization; 2000.
12. Kleynhans EH. Spatial and temporal distribution of trace elements in aerosols in the Vaal triangle [Msc thesis]. Potchefstroom: North-West University; 2008.
13. Heal MR, Hibbs LR, Agius RM, Beverland IJ. Total and water-soluble trace metal content of urban background PM_{10}, $PM_{2.5}$, and black smoke in Edinburgh, UK. Atmos Environ. 2005;39(8):1417–1430. http://dx.doi.org/10.1016/j.atmosenv.2004.11.026
14. Engelbrecht JP, Swanepoel L, Chow JC, Watson JG, Egami RT. The comparison of source contributions from residential coal and low-smoke fuels, using CMB modelling in South Africa. Environ Sci Pol. 2002;5:157–167. http://dx.doi.org/10.1016/S1462-9011(02)00029-1
15. Nriagu J, Champak J, Naidoo R, Coutsoudis A. Atmospheric lead pollution in KwaZulu-Natal, South Africa. Sci Total Environ. 1996;191(1–2):69–76. http://dx.doi.org/10.1016/0048-9697(96)05249-7
16. Kgabi NA. Monitoring the levels of toxic metals of atmospheric particulate matter in the Rustenburg district. Potchefstroom: North-West University; 2006.
17. Baker PGL, Brunke E-G, Slemr F, Crouch AM. Atmospheric mercury measurements at Cape Point, South Africa. Atmos Environ. 2002;36(14):2459–2465. http://dx.doi.org/10.1016/S1352-2310(01)00531-3
18. Venter AD, Vakkari V, Beukes JP, Van Zyl PG, Laakso H, Mabaso D, et al. An air quality assessment in the industrialised western Bushveld Igneous Complex, South Africa. S Afr J Sci. 2012;108(9/10), Art. #1059, 10 pages. http://dx.doi.org/10.4102/sajs.v108i9/10.1059
19. Cramer LA, Basson J, Nelson LR. The impact of platinum production from UG2 ore on ferrochrome production in South Africa. J S Afr I Min Metall. 2004;104(9):517–527.
20. Xiao Z, Laplante AR. Characterizing and recovering the platinum group minerals – a review. Miner Eng. 2004;17:961–979. http://dx.doi.org/10.1016/j.mineng.2004.04.001
21. Beukes JP, Dawson NF, Van Zyl PG. Theoretical and practical aspects of Cr(VI) in the South African FeCr industry. J S Afr I Min Metall. 2010;110:743–750.

22. Department of Home Affairs. Notice 495 of 2012. National Environmental Management: Air Quality Act of 2004, Declaration of the Waterberg National Priority Area, South African Government Gazette No. 35345 on 15 June 2012; Correction notice (154): Waterberg-Bojanala National Priority Area, South African Government Gazette No. 36207 on 8 March 2013.

23. Kaonga B, Kgabi NA. Atmospheric particulate matter in the Marikana mining area of Rustenburg, South Africa. Eur J Sci Res. 2009;34:271–279.

24. Piketh S, Van Nierop M, Rautenbach C. Rustenburg local municipality air quality management plan. Witbank: Palace Consulting Engineers Ltd.; 2005.

25. Laakso L, Laakso H, Aalto PP, Keronen P, Petäjä T, Nieminen T, et al. Basic characteristics of atmospheric particles, trace gases and meteorology in a relatively clean southern African savannah environment. Atmos Chem Phys. 2008;8:4823–4839. http://dx.doi.org/10.5194/acp-8-4823-2008

26. Petäjä T, Vakkari V, Pohja T, Nieminen T, Laakso H, Aalto PP, et al. Transportable aerosol characterization trailer with trace gas chemistry: Design, instruments and verification. Aerosol Air Qual Res. 2013;13:421–435.

27. Vakkari V, Laakso H, Kulmala M, Laaksonen A, Mabaso D, Molefe M, et al. New particle formation events in semi-clean South African savannah. Atmos Chem Phys. 2011;11:3333–3346. http://dx.doi.org/10.5194/acp-11-3333-2011

28. Mouli PC, Mohan SV, Balaram V, Kumar MP, Reddy SJ. A study on trace elemental composition of atmospheric aerosols at a semi-arid urban site using ICP-MS technique. Atmos Environ. 2006;40:136–146. http://dx.doi.org/10.1016/j.atmosenv.2005.09.028

29. United States Environmental Protection Agency (EPA). Compendium methods for the determination of inorganic compounds in ambient air. Compendium method IO-3.5: Determination of metals in ambient particulate matter using inductively coupled plasma / mass spectrometry (ICP/MS). Washington DC: EPA; 1999.

30. Polissar AV, Hopke PK, Paatero P, Malm WC, Sisler JF. Atmospheric aerosol over Alaska 2: Elemental composition and sources. J Geophys Res D. 1998;103(D15):19045–19057.

31. Koutrakis P, Spengler JD. Source apportionment of ambient particles in Steubenville, OH using specific rotation factor analysis. Atmos Environ. 1987;21(7):1511–1519. http://dx.doi.org/10.1016/0004-6981(87)90313-1

32. Brown RJC, Butterfield DM, Goddard SL, Muhunthan D, Brown AS, Beccaceci S, et al. Report to the Department of Environment, Food and Rural Affairs, and the Devolved Administrations, by the National Physical Laboratory: Annual Report for 2009 on the UK Heavy Metals Monitoring Network. Teddington: National Physical Laboratory; 2010.

33. Vieira BJ, Biegalski SR, Freitas MC, Landsberger S. Atmospheric trace metal characterisation in industrial area of Lisbon, Portugal. J Radioanal Nucl Chem. 2006;270(1):55–62. http://dx.doi.org/10.1007/s10967-006-0309-0

34. Bilos C, Colombo JC, Skorupka CN, Rodrigues Presa MJ. Sources, distribution and variability of airborne trace metals in La Plata City area, Argentina. Environ Pollut. 2001;111:149–158. http://dx.doi.org/10.1016/S0269-7491(99)00328-0

35. European Commission. Air quality standards [homepage on the Internet]. No date [updated 2012 Nov 30; cited 2013 Jan 24]. Available from: http://ec.europa.eu/environment/air/quality/standards.htm.

36. United States Environmental Protection Agency. National Ambient Air Quality Standards (NAAQS) [homepage on the Internet]. No date [updated 2012 Dec 14; cited 2013 Jan 24]. Available from: http://www.epa.gov/air/criteria.html.

37. National Environmental Management Air Quality Act of 2004 section 63. Government Gazette Republic of South Africa. 2009,534:8–9. c2004 [cited 2013 Jan 24]. Available from: www.info.gov.za.

38. Lin C-J, Pehkonen SO. The chemistry of atmospheric mercury: A review. Atmos Environ. 1999;33:2067–2079. http://dx.doi.org/10.1016/S1352-2310(98)00387-2

39. Witt MLI, Meheran N, Mather TA, De Hoog JCM, Pyle DM. Aerosol trace metals, particle morphology and total gaseous mercury in the atmosphere of Oxford, UK. Atmos Environ. 2010;44:1524–1538. http://dx.doi.org/10.1016/j.atmosenv.2010.01.008

40. Makkonen U, Hellén H, Anttila P, Ferm M. Size distribution and chemical composition of airborne particles in south-eastern Finland during different seasons and wildfire episodes in 2006. Sci Total Environ. 2010;408:644–651.

41. Scorgie Y. Urban air quality management and planning in South Africa [PhD thesis]. Johannesburg: University of Johannesburg; 2012.

42. Kemink M. A holistic environmental approach to the processing of off gas wastes arising from ferro manganese alloy production [MTech thesis]. Johannesburg: Technikon Witwatersrand; 2000.

43. Allen AG, Nemitz E, Shi JP, Harrison RM, Greenwood JC. Size distribution of trace metals in atmospheric aerosols in the United Kingdom. Atmos Environ. 2001;35:4581–4591. http://dx.doi.org/10.1016/S1352-2310(01)00190-X

44. Heuglin C, Gehrig R, Baltensperger U, Gysel M, Monn C, Vonmont H. Chemical characterisation of PM2.5, PM10 and coarse particles at urban, near-city and rural sites in Switzerland. Atmos Environ. 2005;39:637–651. http://dx.doi.org/10.1016/j.atmosenv.2004.10.027

45. Beukes JP, Pienaar JJ, Lachmann G, Giesekke EW. The reduction of hexavalent chromium by sulphite in wastewater. Water SA. 1999;52:363–370.

46. Hosiokangas J, Ruuskanen J, Pekkanen J. Effects of soil dust episode and mixed fuel sources on source apportionment of PM_{10} particles in Kuopio, Finland. Atmos Environ. 1999;33:3821–3829. http://dx.doi.org/10.1016/S1352-2310(98)00400-2

47. Kennedy BA. Surface mining. Baltimore, MD: Port City Press Inc.; 1990.

An evaluation for harnessing low-enthalpy geothermal energy in the Limpopo Province, South Africa

AUTHORS:
Taufeeq Dhansay[1,2]
Maarten de Wit[2]
Anthony Patt[3,4]

AFFILIATIONS:
[1]Council for Geoscience, Polokwane, South Africa
[2]AEON, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa
[3]International Institute for Applied Systems Analysis, Laxenburg, Austria
[4]ETH, Zürich, Switzerland

CORRESPONDENCE TO:
Taufeeq Dhansay

EMAIL:
taufeeq.dhansay@gmail.com

POSTAL ADDRESS:
PO Box 620, Polokwane 0700, South Africa

DATES:
Received: 06 Sep. 2013
Revised: 01 Nov. 2013
Accepted: 05 Nov. 2013

KEYWORDS:
climate change; renewable energy; geothermal energy; enhanced geothermal systems; energy policy

HOW TO CITE:
Dhansay T, De Wit M, Patt A. An evaluation for harnessing low-enthalpy geothermal energy in the Limpopo Province, South Africa. S Afr J Sci. 2014;110(3/4), Art. #2013-0282, 10 pages. http://dx.doi.org/10.1590/sajs.2014/20130282

South Africa generates most of its energy requirements from coal, and is now the leading carbon emitter in Africa, and has one of the highest rates of emissions of all nations in the world. In an attempt to decrease its CO_2 emissions, South Africa continues to research and develop alternative forms of energy, expand on the development of nuclear and has began to explore potentially vast shale gas reserves. In this mix, geothermal has not been considered to date as an alternative energy source. This omission appears to stem largely from the popular belief that South Africa is tectonically too stable. In this study, we investigated low-enthalpy geothermal energy from one of a number of anomalously elevated heat flow regions in South Africa. Here, we consider a 75-MW enhanced geothermal systems plant in the Limpopo Province, sustainable over a 30-year period. All parameters were inculcated within a levelised cost of electricity model that calculates the single unit cost of electricity and tests its viability and potential impact toward South Africa's future energy security and CO_2 reduction. The cost of electricity produced is estimated at 14 USc/KWh, almost double that of coal-generated energy. However, a USD25/MWh renewable energy tax incentive has the potential of making enhanced geothermal systems comparable with other renewable energy sources. It also has the potential of CO_2 mitigation by up to 1.5 gCO_2/KWh. Considering the aggressive nature of the global climate change combat and South Africa's need for a larger renewable energy base, low-enthalpy geothermal energy could potentially form another energy option in South Africa's alternative energy basket.

Introduction

As the leading carbon emissive nation in Africa, South Africa joined several developing nations with the aim to decrease their greenhouse gas (GHG) emissions. Previously, only developed nations had stringent environmental policies toward decreasing their GHG emissions; however, this situation will change as the rate of GHG emissions from developing nations are predicted to surpass that of developed nations before 2020.[1] Whilst the Kyoto Protocol proposed financial penalties on ratified developed nations, developing nations were given temporary exemption. However, newer planned international legislation is expected to penalise developing nations for failing to meet millennium GHG reduction targets. This legislation was decreed under the Bali Action Plan in which key mitigation scenarios were outlined for developing nations at the COP17 summit held in Durban, South Africa.[2]

Having ratified the Kyoto Protocol, the South African government affirmed its commitment to reduce CO_2 emissions by introducing environmental legislation that includes the *Environmental Act, 1998*[3] and the White Paper on Renewable Energy, 2003[4]. These policy documents highlight the plan to reduce GHG emissions by 40% by 2050 through the development and implementation of a basket of renewable energy sources, including wind, hydro, bio and extensive solar projects. However, with continuous pressure on the South African energy sector to meet immediate and an ever-increasing energy demand, the development of two large-scale coal-fired power plants was commissioned. These plants will not only double the country's energy capacity, but also increase South Africa's overall GHG emissions. Therefore, the current renewable energy scheme will not be sufficient to adequately decrease South Africa's overall GHG emissions without severe market penalties and trade solutions. The National Treasury therefore proposed the introduction of a carbon tax from 2015[5] which, it has been argued, will be potentially damaging to the national economy and forego socio-economic benefits and increasing unemployment[6]. In order to achieve millennium targets and reach a sustainable future, South Africa will need to expand the current renewable energy research protocol and consider other possible alternative energy sources.

Geologically, South Africa is largely underlain by the Archean-aged (greater than 2.5 billion years old) Kaapvaal Craton. This craton comprises a crust of some of the oldest rocks on Earth that formed a nucleus for later continental growth by accreting to it younger rock formations, predominantly between ca. 1–2 billion years old (known as the Namaqua-Natal Mobile Belt).[7,8] The unique stability of the Kaapvaal Craton is because its crust is underlain by an equally old, exceptionally thick and chemically depleted mantle lithosphere with low heat conductivity. This mantle keel reaches depths of ca. 250–300 km,[9-11] and is relatively immune to melting. Together with its overlying crust, this unusually thick and stable lithosphere essentially acts as an insulator, deflecting most of the heat from the underlying convective mantle away from the craton toward younger surrounding regions with less depleted and thinner mantle lithosphere (<120 km).[12,13] These cratonic aspects of the South African geology are the fundamental cause of the low heat flow values measured at the surface in much of the central parts of the country, resulting in low geothermal gradients, which in turn are the fundamental reasons why geothermal energy is not being considered in South Africa as a viable option.[14] Globally, there are a number of countries with similar geological profiles to South Africa (e.g. Canada, Russia and Australia) that have dismissed the geothermal energy potential on their cratonic regions. However, recent experience has shown that these areas may still have capacity to yield energy from relatively low temperatures of 100–200 °C at a depth ranging from 2 km to 3 km[15] – conditions that are met in other parts of South Africa. These areas then provide potential for sustainable low-enthalpy geothermal energy extraction.

We investigated low-enthalpy geothermal energy as a viable alternative energy source in South Africa, by considering a hypothetical geothermal energy plant in the Makuleni Village within the Limpopo Belt. This target area reveals the presence of a suitable heat flow profile, as witnessed by hot springs with surface temperatures of up to 70 °C.[16-18] We present a model that considers the development of a 75-MW enhanced geothermal systems plant, sustainable over a period of 30 years in this area. Attaining an economic viability indication is achieved through levelised cost of electricity (LCOE) model calculations. This model considers all financial and economic parameters associated with the building and maintenance of such a plant, weighed against all geological and engineering parameters associated with harnessing the energy. Because there are other regions in South Africa (notably in the Northern Cape) that have similarly high heat flow profiles, the results of this study should generate recommendations for the consideration of geothermal energy in South Africa's shift toward long-term renewable energy production.

Enhanced geothermal systems

Enhanced geothermal systems (EGS) are unlike hydrothermal or volcanic-fuelled systems that derive heat directly from the Earth's convective mantle. An EGS exploits latent heat encapsulated in deep crustal rocks. These crustal rocks are predominately granitic in composition and have high concentrations of heat producing elements that emit heat during their radioactive decay. These elements include U, Th and K.[19] High heat producing granites located at the Cooper Basin EGS project, South Australia, exhibit values of ca. 30 ppm U and 144 ppm Th, giving rise to temperatures between ca. 250 °C and 350 °C at depths of ca. 4–5 km.[20] This project indicates the possibility of harnessing EGS in South Africa where similar concentrations of heat-producing elements are present in the Namaqualand, Northern Cape and the Limpopo Province.

An EGS system functions by pumping a 'working fluid' (mainly water) into a geothermal reservoir and allowing it to interact with the surrounding hot rocks and heat up substantially. The heated fluid is then pumped back to the surface and used in a binary generation system (Figure 1). The binary system has a second organic-based fluid with a much lower boiling point, which flashes to steam upon exposure to the heated geothermal fluid. The organic steam is then used to run a generator and produce electricity.[21] EGS is also being proposed as a mechanism for simultaneous carbon sequestration, which is thought to be possible by incorporating CO_2 into the working fluid. The CO_2 would theoretically be trapped and stored in the reservoir before the working fluid returns up the production wells.[22]

A successful EGS plant requires an appropriate crustal lithology that contains a relatively high concentration of radioactive heat-producing

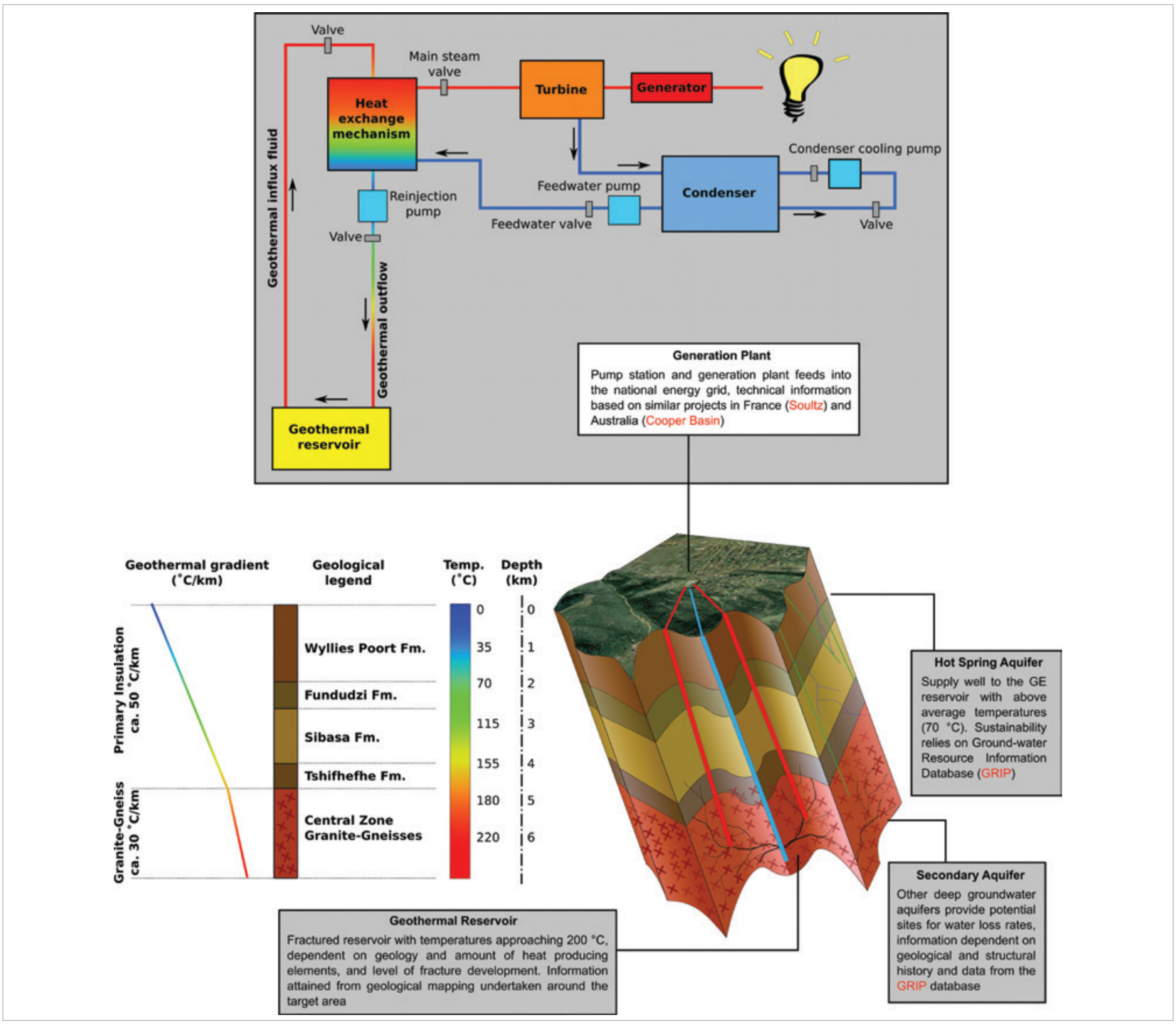


Figure 1: Schematic illustration of the hypothetical enhanced geothermal system plant in the Limpopo Province. The blue line indicates the influx well and the red line indicates the outflow/production wells. Also shown is the depth, geological profile, geothermal gradient and detailed binary production system.

elements. This lithology also needs to attain an appropriate thickness and depth and have a relatively homogenous mineral composition. The crustal lithology will also need to be effectively insulated, by an adequately thick overlying sedimentary sequence. In order to maintain efficient flow between the inflow and outflow systems, a sufficiently porous geothermal reservoir is needed. Therefore, the crustal lithology should have a uniform fracture network that can be manipulated through hydraulic fracturing. Manipulation includes increasing the porosity of the fracture network and creating a sizable geothermal reservoir. Knowledge of the geological mineral chemistry is important to avoid possible decreases in the reservoir porosity as a result of secondary mineral precipitation.

Finding a target area in South Africa

Early heat flow measurements in South Africa have indicated the presence of a relatively low heat flow profile.[23-25] However, these measurements were made within the central part of the Kaapvaal Craton, across the Witwatersrand Basin of the Gauteng region, and reflect the thermal insulating effects of the underlying depleted mantle keel. When heat flow measurements were extended beyond the Witwatersrand Basin, off the Kaapvaal Craton and across the surrounding Proterozoic mobile belts, the heat flow profile increased markedly.[26,27] Measurements show that the heat flow increased from ca. 45 MW/m^2 on the Kaapvaal Craton to ca. 80 MW/m^2 along the Namaqua Natal Mobile Belt.[28,29] This finding suggests that the mantle heat deflected by the cratonic keel may be concentrated beneath thinner neighbouring Proterozoic mobile belts.[12] The higher heat flow evident along the Proterozoic mobile belts can also be attributed to the difference in geology between these terrains. Archean-aged granites are predominantly tonalitic in composition and lack very high concentrations of the heat-producing elements that are found within the younger Proterozoic granites. Figure 2 illustrates a map of South Africa compounding all available heat flow measurements. This map best illustrates the effect of the Kaapvaal Craton with low heat flow signatures and the surrounding Proterozoic mobile belts, which display much higher signatures. This observation suggests that while the Kaapvaal Craton has a very low geothermal potential, the surrounding Kalahari Shield, which includes Proterozoic mobile belts, has a much higher potential.

An important factor in locating a possible EGS target site is the consideration of the concentration of heat-producing elements. Andreoli et al.[30] investigated the gamma ray exposure signatures across South Africa. These signatures highlight regions of South Africa with high concentrations of U and Th (and K), such as in the Proterozoic granites. Many of these signatures correspond to Proterozoic mobile belts of the Kalahari Shield, including the Namaqua Natal and Limpopo Belts.

Careful consideration must also be given to the effect on the surrounding groundwater supply, because EGS uses a substantially larger volume of water than other renewable energy sources. Controlling the amount of groundwater required can be addressed by attempting to create an ideally sealed geothermal reservoir; creation and modification of a fractured reservoir, in turn, necessitates very precise hydraulic fracturing. Early EGS projects in Japan experienced water loss rates of up to 50% because of imprecise fracture modelling and uncontrollable reservoir growth.[31]

A target site in the Limpopo Belt

Given the abovementioned constraints, we considered a target site close to the Makuleni Village within the largely rural Vhembe District of the Limpopo Province (Figure 3). Makuleni and its neighbouring villages have the additional benefit of being located on the national energy grid and included in Eskom's (South Africa's public electricity utility) basic free energy policy, which allocates each household 50 KWh of free electricity a month.

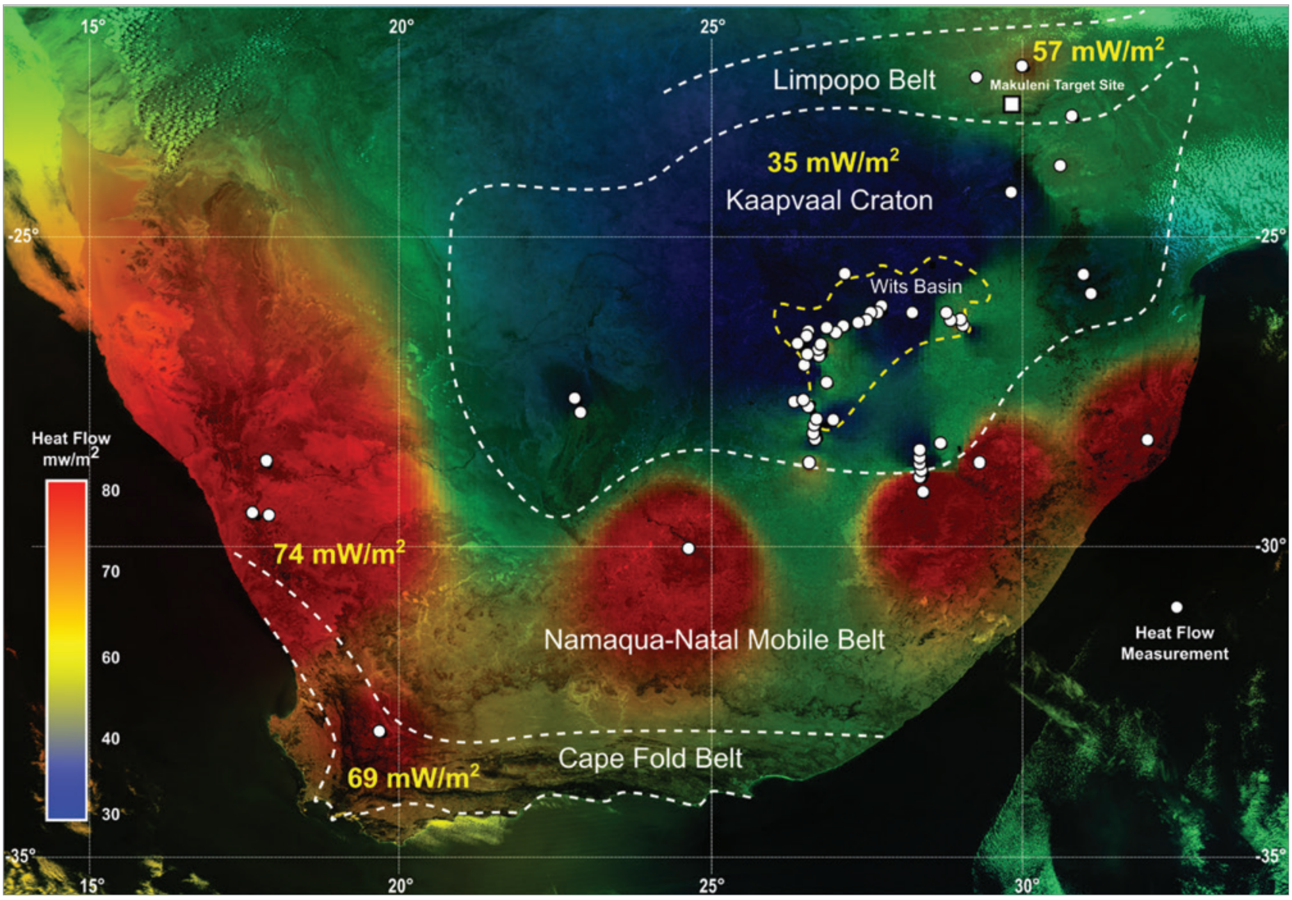


Figure 2: Geothermal potential map produced using available heat flow data created using inverse distance weighting within Quantum GIS. Major tectonic boundaries and geological features are highlighted.

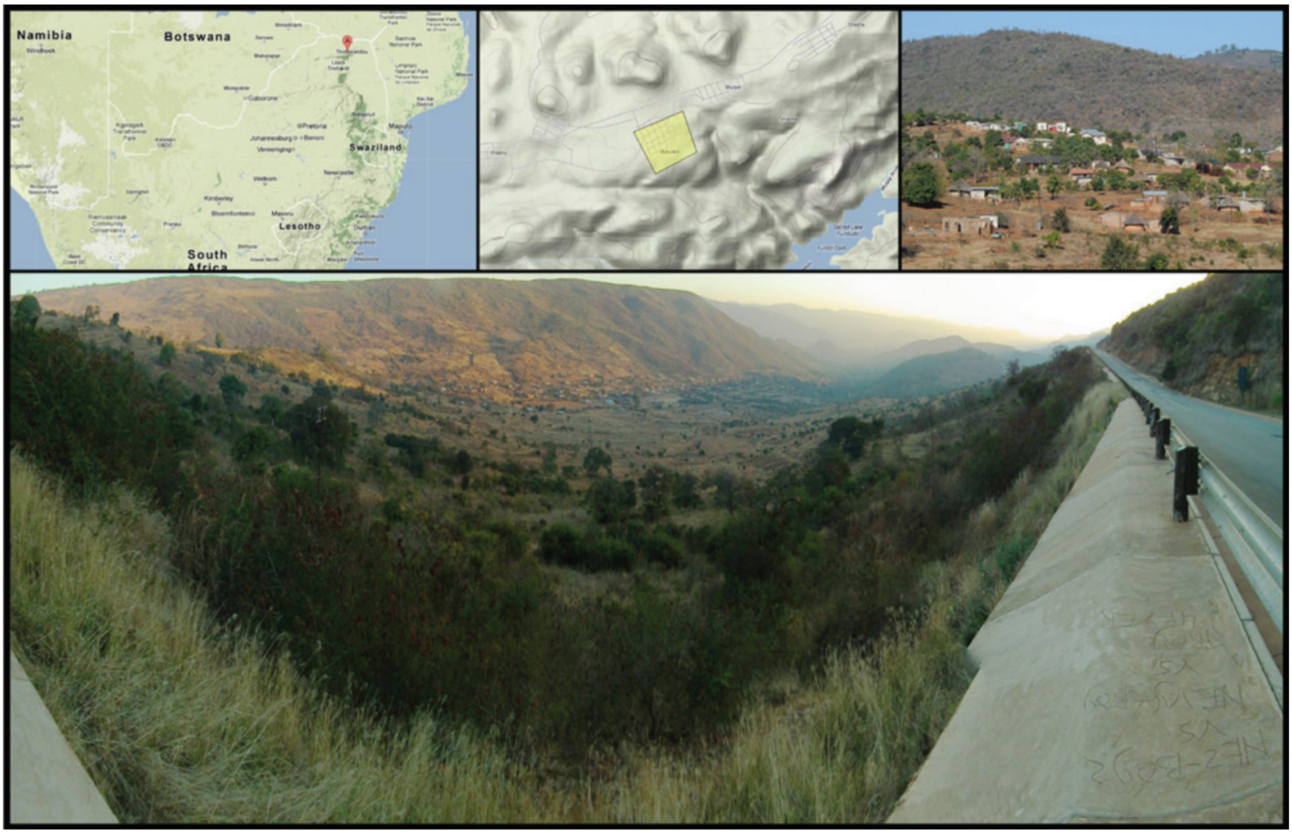


Figure 3: Overview of the Siloam Valley and Makuleni Village: the site of the hypothetical enhanced geothermal system.

The target site is located within the Limpopo Belt, a ca. 150-km long orogenic belt formed during the amalgamation of the Kaapvaal and Zimbabwe cratons. This major period of continental collision began ca. 3.2 Ga, reaching its maximum compressional regime at ca. 2.0 Ga.[32] Outpouring of granites rich in heat-producing elements are evident of this period, providing much of the high heat flow signatures observed.[30,33] The Limpopo Belt is separated into three discrete zones: the northern and southern marginal zones and the central zone, each of which is bounded by major shear features. The marginal zones are largely composed of fragments of their respective adjacent cratons (Zimbabwe and Kaapvaal cratonic fragments in the northern and southern marginal zones, respectively), while the central zone has a unique lithological array.

The Makuleni village is located near the boundary between the central and southern marginal zones, ca. 20 km north of the Kaapvaal Craton boundary. The target site sits on the 1.85 Ga volcano-sedimentary Soutpansberg Group[34,35], which covers the tectonic boundary between the Limpopo Belt and the Kaapvaal Craton and unconformably overlies Limpopo Archean and Palaeoproterozoic granitic rocks[7,36-38]. The Soutpansberg Group was deposited as a syn-rift sequence, in a half-graben basin structure with alternating phases of sedimentary deposition and volcanic outpourings.[39-41] This deposition occurred during a late-stage period of regional extension, following the collision of the Kaapvaal and Zimbabwe cratons.[42] The Soutpansberg Group displays a complex network of normal faults and numerous dolerite dykes and sills that characterise the area with many fault-controlled hot springs. These hot springs further illustrate the higher heat flow of this region. The highest recorded hot spring is found near the Makuleni target site, and reaches a surface temperature of 70 °C, after circulating to a maximum depth of 2 km (based on geological investigations). A possible basal reservoir temperature capable of sustaining low-enthalpy EGS production could be attained at greater depths.

Model development for determining EGS viability

To determine the possible economic viability of harnessing low-enthalpy geothermal energy in the Limpopo region, a LCOE model was developed (Figure 4).[43] This model considers all economic and financial factors associated with building a geothermal energy plant in the Makuleni Village. These factors are then weighed against the total possible energy capable of being generated from the geothermal energy plant. The hypothetical EGS plant in the Makuleni Village is modelled based on similar plants in France[44] and Australia[45].

Model calculations

Levelised cost of electricity

The LCOE model calculations append all cost factors with the potential energy factors to estimate a unit cost of electricity. This calculation is commonly used in most energy producing technologies.[46] This is the final calculation within the model sequence and is expressed as:

$$LCOE = \frac{TC}{TE} \qquad \text{Equation 1}$$

where TC is the total cost and TE is the total energy.

Generation cost

The generation cost considers the total capital cost associated with an EGS plant in the Limpopo Province. This calculation includes the cost of generating turbines and grid parity for feeding back into the national energy grid. Deep drilling and hydraulic fracturing costs have also been incorporated. The calculation is:

$$C(MW) = \frac{C}{N} \qquad \text{Equation 2}$$

where C(MW) is the capital cost per MW of electricity produced, C is the total capital cost and N is the nameplate plant size.

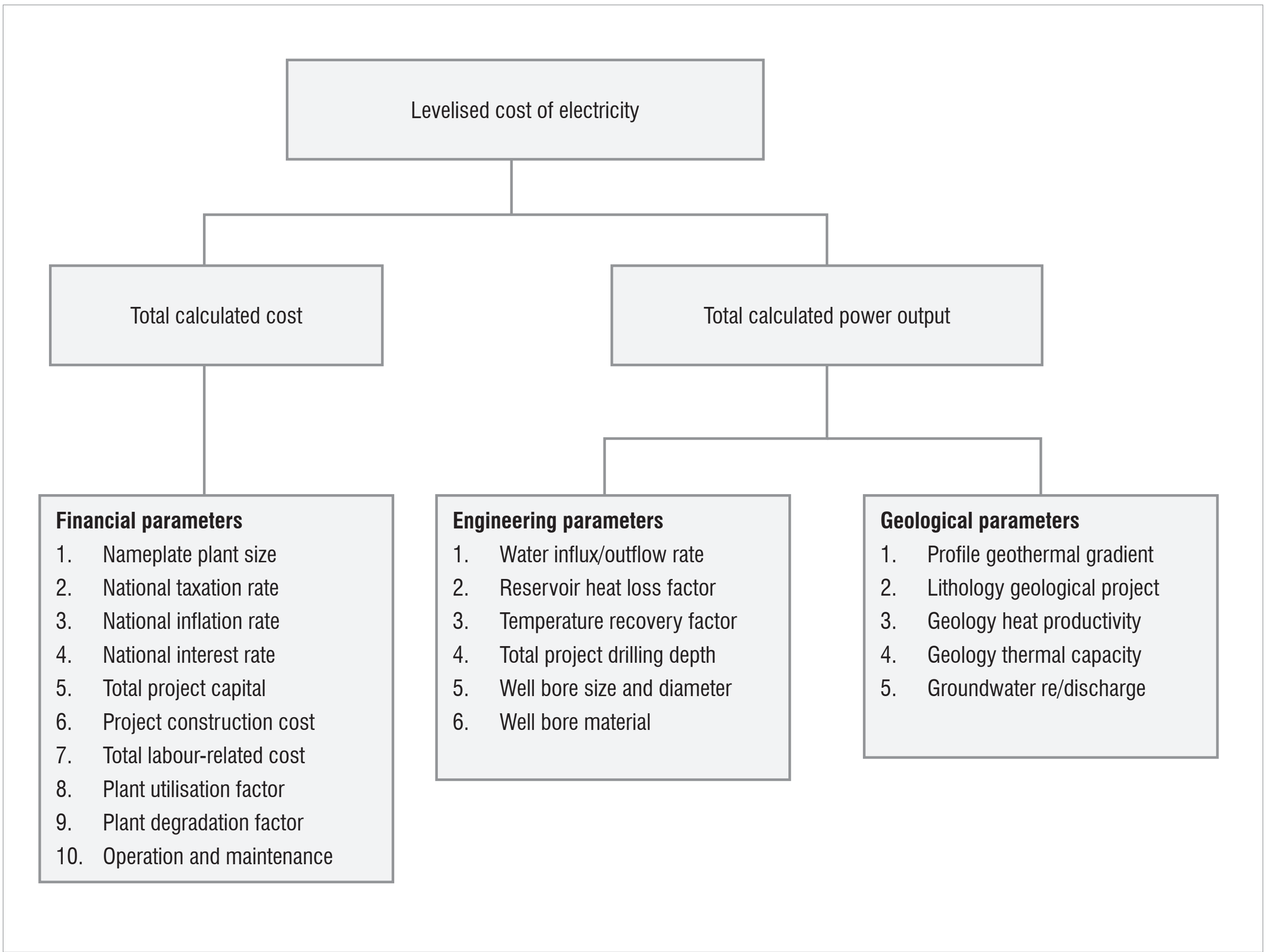


Figure 4: Flow diagram of the economic model developed within this study for the hypothetical enhanced geothermal system.

Operating and maintenance costs

The operating and maintenance costs relate to the various parameters associated with all maintenance-related events. This amount is calculated using a defined operational and maintenance cost per year, weighed against the total energy production of that same year. Both South African tax and inflation rates are included.

$$OM = \sum_{t=2}^{t=n} OM(t) * (1 + i) * P(t) * (1 - T) \quad \text{Equation 3}$$

OM is the overall operational and maintenance costs, t is the year; n is the lifetime of the plant, OM(t) is the defined operational and maintenance value per MWh, i is the South African annual inflation rate, P(t) is the annual electricity production and T is the South African tax rate.

Renewable energy production incentive factor

Renewable energy feed-in tariff (REFIT) is a mechanism used to accelerate the development and implementation of renewable energy. Various industrialised countries have successfully introduced REFIT schemes to assist in the production of renewable energy. Similarly, the National Energy Regulator of South Africa introduced a REFIT scheme that supports various renewable energy projects. While this scheme does not include geothermal energy, this model will consider a geothermal REFIT as follows:

$$PTC = PTC(t) * E(t) * 1000 \quad \text{Equation 4}$$

where PTC is the overall production tax credit, PTC(t) is the defined yearly production tax credit and E(t) is the yearly electricity production.

Geological parameters

The geological parameters are used together with the engineering parameters to determine the most probable energy production. The geology is dependent on the lithological profile underlying the target site and extends until the maximum basal reservoir depth. This parameter determines the variability in heat productivity, conductivity and the overall geothermal gradient along the well length. This calculation is expressed as follows:

$$Tb = \sum_{t=2}^{t=n} \frac{(G * t)}{d} * (1 - L) \quad \text{Equation 5}$$

where Tb is the expected basal reservoir temperature, l is the lithology, n is the number of lithological units, G is the geothermal gradient, t is the thickness of the lithological unit, d is the well depth and L is the thermal loss coefficient.

Engineering parameters

The engineering parameters are used to consider the amount of heat available for exploitation and to estimate the total amount of electricity capable of being harnessed. These parameters include the initial and final temperature of the working fluid, together with the heat capacity and overall flow rate within the system. Variation in the composition of the working fluid can play a major role in the overall energy production.

Here the model assumes the use of water as the main geothermal working fluid. Experiments do suggest that supercritical CO_2 has a lower boiling point and could increase the system efficiency, while allowing for simultaneous carbon sequestration. These factors are equated as follows:

$$Energy = [R * 1000 * C((Tb * (1 - Rf)) - Ti)] * 0.001 \qquad \text{Equation 6}$$

where R is the system flow rate, C is the heat capacity factor, Ti is the input temperature of the working fluid, Tb is the reservoir temperature and Rf is the final recovery factor.

Groundwater sustainability

The process by which geothermal energy is harnessed requires a large volume of water for operation. Failure to create an adequately sealed reservoir can result in poor water recovery rates, which will require the continuous addition of water. The model estimates the value of water required to run the proposed EGS plant and weighs the estimate against the current local exploitation values, together with the natural recharge and discharge rates. A population coefficient is also taken into account and considers an increase in the surrounding population size. An additional coefficient is also equated for any possible water loss. These parameters are equated as follows:

$$W = \sum_{t=1}^{t=n} [Rn - Ry - Gl - Dn - (Dp *) 1 + ip))] \qquad \text{Equation 6}$$

where W is estimated required water supply, t is the year of n years, n is the lifetime of the plant, Rn is the natural recharge rate, Ry is the yearly plant water flow rate required, Gl is the geothermal water loss coefficient, Dn is the human population required discharge and ip is the human population increase factor over the lifetime of the plant.

Results

Levelised cost of electricity

Modelled results compared against other models

The LCOE modelled results of a proposed EGS plant within the Makuleni target site are shown in Figure 5 and are compared with the LCOE of other international EGS projects. The first result, LCOE, is the modelled result for a scenario excluding any REFIT. This scenario would best simulate an EGS plant under the current energy scheme in South Africa that lacks any financial incentives for geothermal energy. LCOE (+REFIT) follows the same parameters, but includes a REFIT of USD25/MWh. This amount is an upper-level accepted geothermal energy incentive used in many international countries.

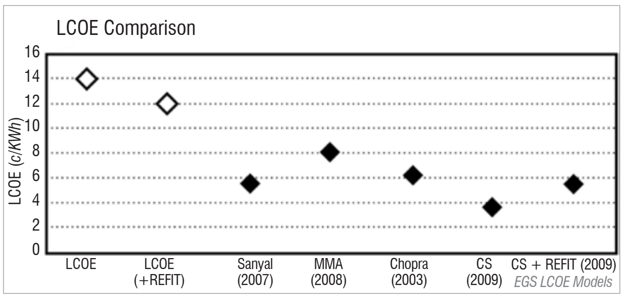
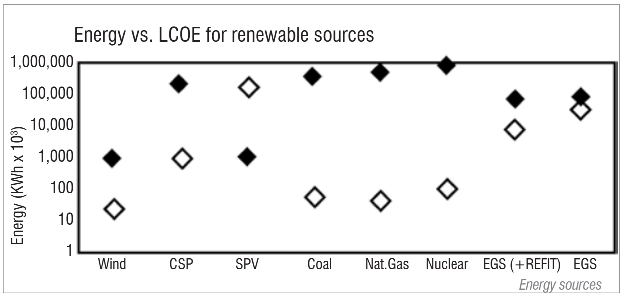


Figure 5: Graph comparing the calculated levelised cost of electricity (LCOE) of this study (white points) with the LCOE of other enhanced geothermal systems (black points).

In Figure 5, Sanyal (2007)[47] represents the modelled LCOE results from the Desert Peak EGS plant in Nevada, USA; MMA (2008)[48] represents the modelled calculations performed by MacLennan Megasanik Associates for the viability of the developing EGS plants in Australia; and Chopra (2003)[49] represents the results as determined for the first successful EGS project in the Cooper Basin, South Australia. The final two points illustrated are modelled results determined by a Credit Suisse report on renewable energy; the first point excludes and the final point includes a REFIT and carbon tax credit.[46] These points had the benefit of actual field data from EGS test sites. This study, however, estimates a much higher LCOE, primarily as a result of the use of conservative data assumptions where actual data is not available.

Calculated energy potential versus other renewable sources

The calculated LCOE and energy capacity of the hypothetical plant were compared with other energy options, including renewable and non-renewable energy (Figure 6). The white points in Figure 6 represent the LCOE of each form of energy, while the corresponding black points represent the total energy capacity of each energy source. Non-renewable sources are clearly more viable, with a greater capacity and lower cost. EGS has the second largest capacity, after concentrated solar power, but it also has the highest cost of all renewable energy sources.



CSP, concentrated solar power; SPV, solar photovoltaic.

Figure 6: Graph comparing the levelised cost of electricity (LCOE) of enhanced geothermal systems (EGS) with those of other renewable energy sources. Black points correspond to the energy capacity and the white points correspond to the LCOE.

Sensitivity analysis

Energy potential versus depth

The change in the LCOE with an increasing basal reservoir temperature and the consequential increase in the overall energy capacity of the system are shown in Figure 7. The dashed curve represents the scenario excluding any REFIT, while the black curve illustrates a scenario including a USD25/MWh REFIT. This calculation restricts the drilling depth to a maximum of 6 km. The temperature range best estimates the expected basal reservoir temperature as determined within the model calculation.

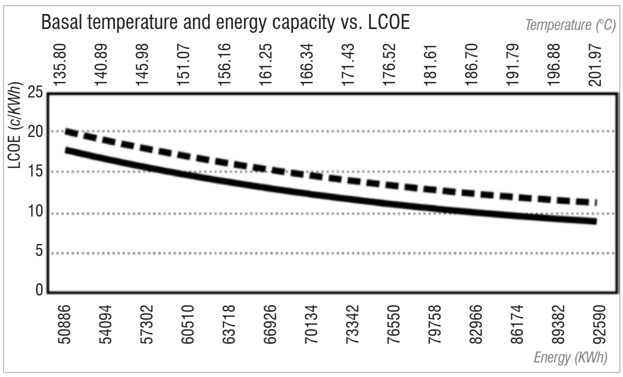


Figure 7: Graph illustrating the effect of an increasing basal reservoir temperature and consequential energy capacity on the levelised cost of electricity (LCOE). The dashed curve represents a scenario excluding a renewable energy feed-in tariff (REFIT), while the solid curve includes a USD25/MWh REFIT.

Cost versus incentives

Renewable energy is an expensive alternative source of energy that relies heavily on governmental incentives to be deemed significantly viable. The effect of a REFIT on the LCOE is shown in Figure 8. The REFIT represents the absolute value of the incentive the government is willing to pay for the production of renewable energy. With no current concession provided for geothermal energy, the effect of an added REFIT ranges from 0 to USD30/MWh of energy produced. The steep decline in the LCOE highlights the need for government-driven renewable energy production incentives.

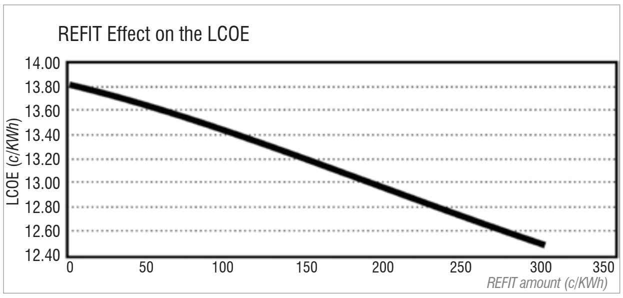


Figure 8: Graph illustrating the effect of an added renewable energy feed-in tariff (REFIT) on the levelised cost of electricity (LCOE) for a hypothetical enhanced geothermal system.

Cost versus drilling depth

The effect of an increasing drilling depth on the LCOE is shown in Figure 9. This scenario considers increasing drilling costs with depth as determined from Augustine[50]. These values are compounded with current and previous inflation rates to obtain a comparable evaluation over the time period specified. This evaluation excludes the additional cost of deep hydraulic fracturing, which is controlled by the geology and fracture transmissivity.

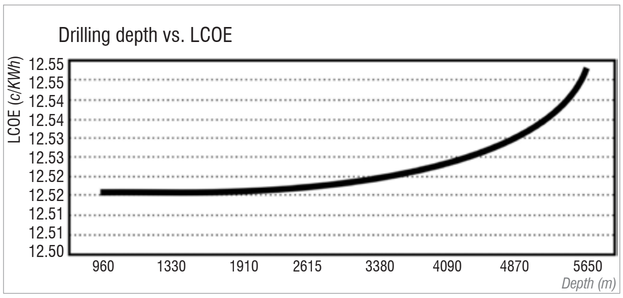


Figure 9: Graph illustrating the effect of an increasing drilling depth on the levelised cost of electricity (LCOE) for a hypothetical enhanced geothermal system.

Sustainability of enhanced geothermal systems

The model calculates the total energy capable of being produced from any production well, at any depth with a corresponding basal reservoir temperature. However, over the lifetime of the plant a thermal drawdown in the basal temperature could result in a decrease in the total energy production. Figure 10 shows the calculated result for the total reduction in heat as indicated by the dashed curve, against the consequential decrease in the overall energy production, indicated by the black curve. The thermal drawdown rate is estimated according to the age and heat productivity of the Limpopo rocks and further appended against those of other EGS granites.

Groundwater sustainability

Water forms the main constituent of the working fluid within the proposed EGS plant system and therefore is an important component within the model. The Makuleni target site is situated where one of the main potable water sources is groundwater. The overall effect of an EGS plant on the groundwater volume is displayed in Figure 11. A large concentration of the water exploited for human disposal is attained from groundwater aquifers stored within fracture zones and incorporated below surface sediment layers. Using data from the Department of Water Affairs of South Africa and the Groundwater Resource Information Project, which monitors several thousand boreholes throughout the Limpopo Province,[51] the model calculates the overall amount of water exploited for human and industrial consumption, and weighs this against the overall recharge rates. Periods of both drought and flooding are also considered. These calculations are forecasted over the lifetime of the EGS plant and are appended with the inclusion of a coefficient for human population size increase associated with further social development within the Limpopo Province. The dashed curve in Figure 11 illustrates the result of this forecast, while the black curve illustrates water consumption of the EGS plant. Based on these results, the idealised closed system of the EGS plant will be sustainable over its lifetime of 30 years.

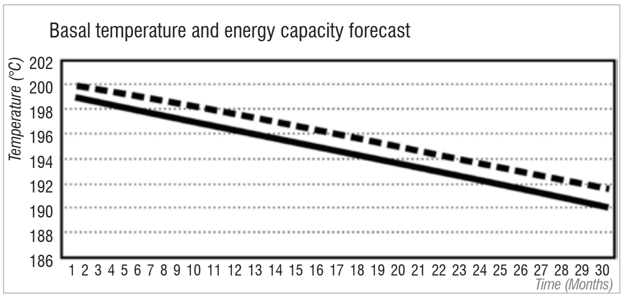


Figure 10: Graph forecasting the decrease in basal reservoir temperature (dashed curve) and the consequential decrease in energy production (solid curve) of a hypothetical enhanced geothermal system.

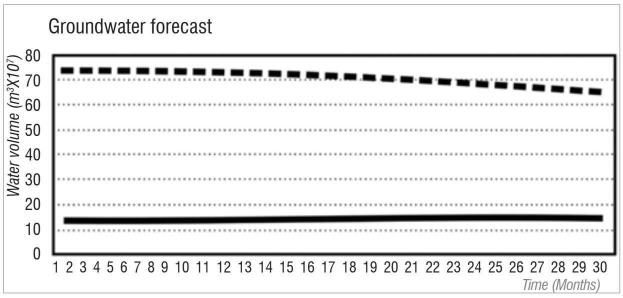


Figure 11: Graph forecasting the groundwater requirements of the hypothetical enhanced geothermal system (solid curve) and the average natural recharge rate (dashed curve).

Discussion

Excluding any REFIT incentive, the EGS LCOE determined the cost of geothermal energy to be greater than 14 USc/KWh in South Africa. This figure is about 7 USc/KWh more than the cost of coal-generated electricity that currently comprises more than 90% of South African electricity. However, under the same circumstances, with an added REFIT of USD25/MWh, the LCOE decreases to approximately 12 USc/KWh. This figure is more comparable with other renewable energy options: about 1 USc/KWh more than solar photovoltaic, although with up to a 25-MW lower energy capacity. When comparing these results against those of other models for similar EGS plants, the results are comparable with the calculated case scenario, deviating no more than 4 USc/KWh from other model projections. As the global and local combat against CO_2 emissions intensifies, nations will face stringent penalties for failing to meet GHG reduction targets. The legal implementation of GHG reduction is becoming more aggressive following the COP18 summit, especially considering the amendment made to the Kyoto Protocol that extends it for another 8-year commitment period. This amendment transpires in addition to a newer climate change agreement expected to be resolved in 2015. The 2015 resolution is expected to have legal implementations that are likely to result in financial penalties for big GHG emitters, like

South Africa. Changes are already present in South Africa with the introduction of the Carbon Tax Policy.[5] This policy aims to reduce GHG emission through the introduction of carbon taxing for large CO_2 emitters in South Africa. Policies like the carbon tax form an integral part of the reduction of South Africa's carbon footprint; however, these policies may potentially have a negative impact on foreign business investment.[6]

The high cost of the EGS plant calculated in this model assumes the prospect of an EGS plant in an area with a predicted lower heat production and heat flow rate than other international target sites. Many of these sites boast some of the highest heat flows in areas lacking active volcanism and/or tectonism. With the lower heat flow rate estimated within the target area, the model is based on the calculations assuming a drilling depth of 5–6 km. Deeper drilling would likely cause a substantial increase to the overall project cost. The model suggests that a continued increase in the drilling depth below 2 km would result in a steeply increasing LCOE, but only of 1–2 USc/KWh, excluding additional hydraulic fracturing costs. Hydraulic fracturing is an additional cost to the deep drilling.

Previous EGS plants in the UK and Japan experienced great fluid losses associated with extended hydraulic fracturing programmes. Hydraulic fracturing was applied in an attempt to control the growth of the reservoir, which often closed because of low fluid flow rates or the accumulation of mineral precipitates. These reports highlight the importance of accurately assessing the orientation and scale of the fractured reservoir and the mineral chemistry of the targeted lithology prior to undertaking hydraulic fracturing. In addition, hydraulic fracturing has the potential to cause irregular surface seismic events. EGS projects in Basal, Switzerland and Landau, Germany were forced to stop following induced seismicity associated with hydraulic fracturing.[52] This further highlights the need for an active seismic-monitoring system during any hydraulic fracturing.

The total energy capable of being produced from an EGS plant within the Limpopo Belt depends predominantly on the basal reservoir temperature encountered. With a shortage of heat flow data, the model relies on information gathered on the surrounding hot springs and granites enriched in radiogenic elements to attain predictions of reservoir temperatures against depth. The calculated basal reservoir temperature illustrates that EGS will be viable and sustainable for a lifespan of 30 years at the Makuleni target site if a basal temperature of 150–200 °C is attained at a depth of 5–6 km. South Africa is well endowed with granites that exhibit large concentrations of heat-producing elements, similar to those in the Cooper Basin, Australia. Considering the high basal reservoir temperatures achieved at the Cooper Basin EGS project, South Africa could have the potential to reach the expected basal reservoir temperatures considered within the model, in a number of regions, notably the rural areas of the North West Province, and likely parts of the Eastern Cape and KwaZulu-Natal Provinces.

Based on the data from the Department of Water Affairs and the Groundwater Resource Information Project in the Limpopo Province, the average rate of aquifer recharge is adequate to sustain an EGS plant in the Makuleni village. Furthermore, the model, based on the average recharge and discharge rates as experienced during possible periods of drought, appended with a coefficient for the increase in demand for groundwater supply over the lifetime of the plant, shows favourable results for sustaining an EGS plant, while having a low impact on the surrounding communities. Assuming a regular thermal decrease in the basal reservoir temperature, the model predicts that there would not be a great drop in the overall production temperature over a period of 30 years, and this drop would amount to less than a 1 MWh decrease in the total energy production.

Conclusion

While geothermal is not currently under consideration by the National Energy Regulator of South Africa, a provision is in place for its inclusion subject to commercial application. Globally, the development of renewable energy is becoming aggressive and the technology of low-enthalpy EGS is advancing toward the use of low-temperature resources. With certain types of organic fluid compositions and specific fluid flow rates, EGS plants can become economically viable at temperatures as low as 100 °C. These conditions could make EGS more viable in many areas around the world.

Several key conclusions and recommendations can be drawn from this work. Firstly, EGS in conjunction with a binary generation system can generate energy from low-enthalpy sources, and has the additional possibility of simultaneously sequestering CO_2. This technology could assist in South Africa's shift toward renewable energy security. Secondly, EGS is a costly form of renewable energy and will rely heavily on governmental tax incentives to achieve viability. Thirdly, South Africa's geology in general, and that of the Kaapvaal Craton in particular, has a relatively low heat flow because of a thick underlying mantle keel. However, there are zones in the broader Kalahari Shield where high heat producing granite-gneisses occur that give rise to much higher heat flow signatures. These areas could form the ideal locations for low-enthalpy EGS exploitation and also correlate with rural communities that could benefit from such development. Finally, if geothermal energy is to be realised in South Africa, further research, especially data acquisition, is required to adequately delineate and design any prospective low-enthalpy EGS plant.

Acknowledgements

This work benefitted from an extended stay at the International Institute for Applied Systems Analysis (IIASA); we are grateful to Robert Kriger of the National Research Foundation for initiating the collaboration with IIASA and providing an impetus for young South African scientists to pursue collaborative projects there in applied science. The work was funded through the Inkaba yeAfrica programme, through the Department of Science and Technology and the National Research Foundation of South Africa. Thanks to the Council for Geoscience for providing facilities and encouragement. A special thank you is also extended to Dr Günther Brandl for his excellent revision of this manuscript. The Inkaba yeAfrica contribution number is 80 and the AEON contribution number is 115.

Authors' contributions

T.D. was the primary researcher with this work forming the basis of his MSc degree at the Nelson Mandela Metropolitan University (NMMU), with a large portion of the work completed at the International Institute for Applied Systems Analysis (IIASA). M.d.W. and A.P. provided supervision and review at NMMU and IIASA, respectively.

References

1. Reid WV, Goldemberg J. Developing countries are combatting climate change: Actions in developing countries that slow growth in carbon emissions. Energ Policy. 1998;26(3):233–237. http://dx.doi.org/10.1016/S0301-4215(97)00137-7

2. United Nations Framework Convention on Climate Change. Report of the Conference of the Parties on its thirteenth session; 2007 Dec 3–15; Bali, Indonesia. Geneva: United Nations; 2008. Available from: http://unfccc.int/documentation/documents/advanced_search/items/6911.php?priref=600004670

3. National Environmental Management Act of 1998 (Act No. 107 of 1998) [statute on the Internet]. c1998 [cited 2014 Mar 03]. Available from: https://www.environment.gov.za/sites/default/files/legislations/nema_amendment_act107.pdf

4. Department of Minerals and Energy, Republic of South Africa. White Paper on Renewable Energy, November 2003 [document on the Internet]. c2003 [cited 2014 Mar 03]. Available from: http://unfccc.int/files/meetings/seminar/application/pdf/sem_sup1_south_africa.pdf

5. Department of National Treasury, Republic of South Africa. Carbon Tax Policy Paper, 2013 [document on the Internet]. c2013 [cited 2014 Mar 03]. Available from: www.saaqis.org.za/filedownload.aspx?fileid=715

6. Lloyd P. Global experience warns SA against imposing carbon tax. Business Day. 2013 August 27; Opinion and Analysis.

7. Stankiewicz J, De Wit M. 3.5 billion years of reshaped Moho, southern Africa. Tectonophysics. 2013;609(8):675–689. http://dx.doi.org/10.1016/j.tecto.2013.08.033

8. Nguuri T, Gore G, James D, Webb S, Wright C, Zengeni T, et al. Crustal structure beneath southern Africa and its implications for the formation and evolution of the Kaapvaal and Zimbabwe cratons. Geophys Res Lett. 2001;28(13):2501–2504. http://dx.doi.org/10.1029/2000GL012587

9. Stankiewicz J, Chevrot S, Van der Hilst RD, De Wit MJ. Crustal thickness, discontinuity depth and upper mantle structure beneath southern Africa: Constraints from body wave conversions. Phys Earth Planet Int. 2002;130(3–4):235–251. http://dx.doi.org/10.1016/S0031-9201(02)00012-2

10. Boyd F, Gurney J. Diamonds and the African lithosphere. Science. 1986;232(4749):472–477. http://dx.doi.org/10.1126/science.232.4749.472

11. Niu F, Levander A, Cooper C, Aeolus C, Lenardic A, James D. Seismic constraints on the depth and composition of the mantle keel beneath the Kaapvaal Craton. Earth Planet Sci Lett. 2004;224(10):337–346. http://dx.doi.org/10.1016/j.epsl.2004.05.011

12. Ballard S, Pollack H. Diversion of heat by Archean cratons: A model for southern Africa. Earth Planet Sci Lett. 1987;85:253–264. http://dx.doi.org/10.1016/0012-821X(87)90036-7

13. Jones MQW. Heat flow and heat production in the Namaqua mobile belt, South Africa. J Geophys Res. 1987;92(B7):6273–6289. http://dx.doi.org/10.1029/JB092iB07p06273

14. Mawson N. South Africa too stable for geothermal. Mail and Guardian. 2013 June 14-20;Supplement p. 44.

15. Axelsson G. Production capacity of geothermal systems. The workshop for decision makers on direct heating use of geothermal resources in Asia; 2008 May 11–18; Tianjin, China.

16. Nyabeze P, Venter J, Olivier J, Motlakeng T. Characterisation of the thermal aquifer associated with the Siloam hot spring in Limpopo, South Africa. In: Totolo O. Proceedings of the 3rd IASTED African Conference on Water Resource Management; 2010 September 6–8; Gaborone, Botswana. Calgary: ACTA Press; 2010. 686-059. http://dx.doi.org/10.2316/P.2010.686-059

17. Olivier J, Van Niekerk H, Van der Walt I. Physical and chemical characteristics of thermal springs in the Waterberg area in the Limpopo Province, South Africa. Water SA. 2008;34(2):163–174.

18. Tekere M, Lotter A, Olivier J, Jonker N, Venter S. Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa. Afr J Biotech. 2011;10(78):18005–18012.

19. Chamberlan CP, Sonder LJ. Heat producing elements and the thermal and baric patterns of metamorphic belts. Science. 1990;250(4982):763–769. http://dx.doi.org/10.1126/science.250.4982.763

20. Uysal T, Gasparon M, Bryan S. QGECE research on delineation of Australian geothermal resources. Australian Geothermal Energy Conference; 2011 November 14–15; Melbourne, Australia.

21. Vogt C, Kosack C, Marquart G. Stochastic inversion of the tracer experiment of the enhanced geothermal system demonstration reservoir in Soultz-sous-Forêts. Revealing pathways and estimating permeability distribution. Geothermics. 2012;42(1):1–12. http://dx.doi.org/10.1016/j.geothermics.2011.11.001

22. Damiani D, Litynski J, McIlvried H, Vikara D, Srivastava R. The US Department of Energy's R&D program to reduce greenhouse gas emissions through beneficial uses of carbon dioxide. Greenhouse Gases: Science and Technology. 2001;2(1):9–16. http://dx.doi.org/10.1002/ghg.35

23. Krige LJ. Borehole temperatures in the Transvaal and Orange Free State. P Roy Soc Lond Mat A. 1939;173(955):450–474. http://dx.doi.org/10.1098/rspa.1939.0158

24. Bouwer RF. Borehole temperatures in the Klerksdorp and Orange Free State areas. Bulletin 22. Pretoria: Geol Sur Union SA Dep Mines; 1954.

25. Carte AE. Thermal conductivity and mineral composition of some Transvaal rocks. Am J Sci. 1955;253(8):482–490. http://dx.doi.org/10.2475/ajs.253.8.482

26. Carte AE, Van Rooyen AI. Further measurements of heat flow in South Africa. Publications of the Geological Society of South Africa, Upper Mantle Project-South African National Committee Symposium. 1969;2(1969):445–448.

27. Chapman DS, Pollack HN. Heat flow and heat production in Zambia: Evidence for lithosphere thinning in central Africa. Tectonophysics. 1977;41(1–3):79–100. http://dx.doi.org/10.1016/0040-1951(77)90181-0

28. Jones MQW. Heat flow anomaly in Lesotho: Implications for the southern boundary of the Kaapvaal Craton. Geophys Res Lett. 1992;19(20):2031–2034. http://dx.doi.org/10.1029/92GL02207

29. Jones MQW. Heat flow in South Africa. Pretoria: Department of Mineral and Energy Affairs; 1992.

30. Andreoli M, Hart R, Ashwal L, Coetzee H. Correlations between U, Th and metamorphic grade in the western Namaqualand Belt, South Africa, with implications for radioactive heating of the crust. J Petrol. 2006;47(6):1095–1118. http://dx.doi.org/10.1093/petrology/egl004

31. Yamaguchi S, Akibayashi S, Rokugawa S, Fujinaga Y, Tenma N, Sato Y. The numerical modelling study of the Hijiori HDR test site. In: Proceedings of the World Geothermal Congress; 2000 May 28–June 10; Kyushu-Tohoku, Japan. Auckland: International Geothermal Association; 2000. p. 3975–3980.

32. Rigby M, Mouri H, Brandl G. A review of the pressure-temperature-time evolution of the Limpopo Belt: Constraints for a tectonic model. J Afr Earth Sci. 2008;50(2008):120–132. http://dx.doi.org/10.1016/j.jafrearsci.2007.09.010

33. Pretorius W. A geochemical and geophysical investigation of a suite of crustal and upper mantle nodules from Venetia kimberlite pipes, Limpopo Belt, South Africa [MSc thesis]. Johannesburg: Rand Afrikaans University; 1996.

34. Cheney ES, Barton JM, Brandl G. Extent and age of the Soutpansberg sequence of South Africa. S Afr J Geol. 1990;93(4):664–675.

35. Barton JM, Pretorius W. Soutpansberg age (1.85 Ga) magmatism and metallogenesis in southern Africa: A result of regional rifting. International Symposium on Plumes, Plates and Mineralisation; 2007 April 12–19; Pretoria, South Africa.

36. Brandl G. The geology of the Pietersburg area: Explanation of sheet 2328. Pretoria: Geological Survey South Africa; 1986.

37. Brandl G. The geology of the Tzaneen area: Explanation of sheet 2330. Pretoria: Geological Survey South Africa; 1987.

38. Brandl G, Kroener A. Preliminary results of single zircon studies from various Archean rocks of north-eastern Transvaal. The 16th Colloqium of African Geology; 1993 September 14–16; Mbabane, Swaziland.

39. Bumby AJ, Eriksson PG, Van der Merwe R, Brummer JJ. Shear-zone controlled basins in the Blouberg area, Northern Province, South Africa: Syn- and post-tectonic sedimentation relating to ca. 2.0 Ga reactivation of the Limpopo Belt. J Afr Earth Sci. 2001;33:445–461. http://dx.doi.org/10.1016/S0899-5362(01)00087-2

40. Crow C, Condie KC. Geochemistry and origin of early Proterozoic volcanic rocks from the Transvaal and Soutpansberg successions, South Africa. Precambrian Res. 1990;47(1–2):17–26. http://dx.doi.org/10.1016/0301-9268(90)90027-N

41. Bumby AJ, Eriksson PG, Van der Merwe R, Steyn GL. A half-graben setting for the Proterozoic Soutpansberg Group (South Africa): Evidence from the Blouberg Area. Sediment Geol. 2002;147:37–56. http://dx.doi.org/10.1016/S0037-0738(01)00186-5

42. Schaller M, Steiner O, Studer I, Holzer L, Herwegh M, Kramers JD. Exhumation of Limpopo central zone granulites and dextral continental-scale transcurrent movement at 2.0 Ga along the Palala Shear Zone, Northern Province, South Africa. Precambrian Res. 1999;96(3):263–288. http://dx.doi.org/10.1016/S0301-9268(99)00015-7

43. Dhansay T. Evaluation for harnessing low-enthalpy geothermal energy in South Africa based on a model pilot plant in the Limpopo Mobile Belt [MSc thesis]. Port Elizabeth: Nelson Mandela Metropolitan University; 2013.

44. Hooijkaas G, Genter A, Dezayes C. Deep-seated geology of the granite intrusions at the Soultz EGS site based on data from 5 km-deep borehole. Geothermics. 2006;35(5):484–506. http://dx.doi.org/10.1016/j.geothermics.2006.03.003

45. Chopra P, Wyborn D. Australia's first hot dry rock geothermal energy extraction project is up and running in granite beneath the Cooper Basin, NE, South Australia. Magmas to Mineralisation: The Ishihara Symposium; 2003 July 22–24; Sydney, Australia.

46. Credit Suisse. How much does it cost? Equity research – Energy technology, alternative energy sector review. 2009 January 28; p.26.

47. Sanyal S, Morrow J, Butler S, Tait A. Cost of electricity from enhanced geothermal systems. Proceedings of the 32nd Workshop on Geothermal Reservoir Engineering; 2007 January 30 – February 01; Stanford, CA, USA. New York: Curran Associates Inc.; 2008.

48. McLennan Megasanik Associates Pty Ltd. Installed capacity and generation from geothermal sources by 2020. Report presented to the Australian Geothermal Energy Association [document on the Internet]. c2008 [cited 2014 Mar 03]. Available from: http://www.infrastructureaustralia.gov.au/public_submissions/published/files/374_agea_SUB3.pdf

49. Chopra P. Status of the geothermal industry in Australia, 2000–2005. Proceedings of the World Geothermal Congress; 2005 Apr 22–29; Antalya, Turkey.

50. Augustine C. A comparison of geothermal with oil and gas well drilling costs. Proceedings of the 31st Workshop on Geothermal Reservoir Engineering; 2006 January 30 – February 01; Stanford, CA, USA. New York: Curran Associates Inc.; 2006.

51. Department of Water Affairs, Republic of South Africa. Ground Water Resource Information Project [homepage on the Internet]. No date [updated 2002 Apr 01; cited 2014 Mar 03]. Available from: http://griplimpopo.co.za/

52. Baisch S, Carbon D, Dannwolf U, Delacou B, Devaux M, Dunand F, et al. Deep heat mining Basel – Seismic risk analysis. Technical report by the SERIANEX Group, presented to the Department fur Wirtschraft, Soziales und Umwelt des Kantos Basel-Stadt, Basel. SERIANEX; 2009.

Predicting scientific research output at the University of KwaZulu-Natal

AUTHOR:
Mike Murray[1]

AFFILIATION:
[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Westville Campus, Durban, South Africa

CORRESPONDENCE TO:
Mike Murray

EMAIL:
murraym@ukzn.ac.za

POSTAL ADDRESS:
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

DATES:
Received: 10 Oct. 2013
Revised: 19 Dec. 2013
Accepted: 01 Jan. 2014

KEYWORDS:
publishing record; zero inflation; Poisson; science faculty; UKZN

HOW TO CITE:
Murray M. Predicting scientific research output at the University of KwaZulu-Natal. S Afr J Sci. 2014;110(3/4), Art. #2013-0321, 4 pages. http://dx.doi.org/10.1590/sajs.2014/20130321

Academic members of staff at the University of KwaZulu-Natal (UKZN) are expected to publish in research journals that have been accredited by the South African based Department of Higher Education and Training. However, some members of staff have chosen to focus solely on the teaching aspect of their careers and as a result they have no publication record. In this study, a set of per annum productivity unit counts was calculated for every academic at UKZN. Because it is possible for a publishing academic to also record a zero count for a given year, it is necessary to develop an appropriate methodology that can distinguish this zero count from one that will always be recorded by a non-publishing academic. By fitting a zero-inflated Poisson model to the data, specific factors can be identified that separately drive the non-publishing and publishing processes at UKZN. In particular, having a PhD and working in a large school has a significant impact on improving the research output of a publishing academic. If UKZN wants to become a research-focused university, non-publishing academics should be encouraged to undertake a PhD degree.

Introduction

In South Africa, the Department of Higher Education and Training allocates funds to universities based on a formula that rewards student throughput as well as the research output of publishing academics.[1,2] In this paper, an attempt is made to identify possible demographic and academic factors that help to improve the research output of academics at the University of KwaZulu-Natal (UKZN). Similar studies have been conducted, mainly on US data (for example Xie and Shauman[3], Aksnes et al.[4] and Kyvik and Teigen[5]), in which it was found that for almost every age group in their respective data sets, men publish more than women. Barjak[6], Gonzalez-Brambila and Veloso[7] and Kyvik[8] have found that research productivity tends to increase with age, reaching a peak before tapering off towards retirement. What distinguishes this study from those mentioned above is that it analyses a set of per annum based publication counts for all members of staff at UKZN. Some staff members will have a zero value for a particular year because they did not publish anything during that year even though they have a record of previous publications, whereas others may have a zero value because they have chosen to focus entirely on the teaching aspect of their careers and therefore never publish. Essentially, a per-annum publication-based output variable is observed, but an accompanying variable denoting whether or not the person recording this zero outcome is (or will eventually become) a publishing academic (or not) cannot be known. A zero-inflation based modelling approach helps to overcome this problem of not being able to identify the true identity of a given record as belonging to that of a publishing or non-publishing academic by introducing into the modelling process an underlying process – process Z – that generates two types of zeroes: a structural zero when dealing with a non-publishing academic and a true zero when dealing with a publishing academic.

Research methodology

A zero-inflation model assumes two possible sources from which a zero observation can arise. Academics who have made a conscious decision to never publish will generate what is called a ‘structural’ zero. Those who have a record of previous publications but have not published in a particular year generate what is called a ‘true’ or ‘sampling’ zero.

Focusing on the structural zero recorded every year for a non-publishing academic, a zero inflation model links the probability p_0 of being able to observe such an outcome with the covariate profile x associated with a particular academic using the following logistic regression model[9]:

$$ln\left(\frac{p_0(\mathrm{x})}{1-Po(\mathrm{x})}\right)=x'\gamma=>p_0\ (\mathrm{x})\ \frac{e^{x'\gamma}}{1+e^{x'\gamma}}$$ Equation 1

Focusing on the per annum based research records generated by a publishing academic for whom, in a particular year, a true zero value may have been recorded, a Poisson distribution with a covariate dependent intensity parameter $\lambda(x)$ that is linked to a set of covariates x using the following function

$$\ln \lambda(x) = x'\theta = > \lambda(x) = e^{x'\theta}$$ Equation 2

is used to model the output generated. With these model choices in hand, a final model formulation – a zero-inflated Poisson model – is obtained for our response variable Y that sets[13]

$$P\ (Y=0)= p_0 + (1-p_0)e^{-\lambda(x)}$$

and

$$P\ (Y=k)= (1-p_0)\ \frac{\theta^{-\lambda(x)}\lambda(x)^k}{k!}$$

for $k=1,2...$

Parameter estimates for ϴ and γ can be obtained using the method of maximum likelihood, details of which can be found in Lambert[9]. Because the mean of a Poisson distribution equals the intensity parameter $\lambda(x)$ that is helping to characterise that distribution, the above expression linking $\lambda(x)$ to the covariates in x implies for any positive estimate that occurs in ϴ that the resulting covariate to which this estimate refers has a positive effect, that is, increases the number of per annum based research outputs produced by that academic at UKZN. Similarly, those covariates that have a negative parameter estimate associated with them are associated with a reduced level of research output.

The response variable

At UKZN, the productivity unit counts that appear in Table 1 have been used to unbiasedly apportion a `unit of worth' to a published piece of work. This approach differs from that of most other studies in which each article that has been published is usually weighted by the journal impact factor which measures the average number of citations a journal paper will receive in the first 2 years following its year of publication. Because the focus in this study was to identify specific factors that affect the production of research rather than on the quality of this research, a 'unit of worth' was chosen to use as a measure, but it can be noted that the method to be outlined can just as easily be applied as an impact factor adjusted response variable. This response variable was then rescaled using the following rule so that it better fitted into the model paradigm outlined in the previous section for a zero-inflated Poisson model:

Y	Number of productivity units achieved
0	No productivity units in that given year
1	1–15 productivity units in that given year
2	16–30 productivity units in that given year
3	31–45 productivity units in that given year
4, 5, etc.	46–60 productivity units, 61–65 units, etc.

In addition to this response variable, the following covariates that help to distinguish one academic from another were also collected:

- a 0/1 indicator variable denoting whether the academic member of staff was female or male
- a set of separate 0/1 indicator variables denoting whether the academic was a lecturer, senior lecturer or professor
- a set of separate 0/1 indicator variables denoting the racial group to which the academic belongs (African, Indian or white)
- a 0/1 indicator variable denoting whether the academic has a PhD or not
- a variable denoting the number of academics in the school in which the academic resides (size)
- an age-based category variable taking on a value 0 if the academic (in that particular year) is in their twenties, a value 1 if they are in their thirties, etc.

Table 1: Productivity unit counts used to record the research output generated by an academic in the Faculty of Science and Agriculture at the University of KwaZulu-Natal

Research activity	Productivity unit count
Journal article† (sole author)	60
Entire book	100
Chapter in book	15
Graduating MSc student	15
Graduating PhD student	60

†Co-authored articles were assigned productivity unit counts on a pro-rata basis; for example, a count of 30 was assigned for an article with two authors.

Results

Data was collected for the period from 2004 to 2008 and consisted of a total of 1236 year-on-year productivity unit counts. A breakdown of the data set according to age, race, gender, qualification and job position is given in Table 2.

Table 2: A demographic breakdown of the academics within the data set for this study

	2004	2005	2006	2007	2008
Age					
20–29	11	9	15	16	13
30–39	60	70	68	67	70
40–49	92	80	87	91	86
50–59	70	74	72	70	71
60+	4	8	11	12	9
PhD					
Yes	195	195	197	198	188
No	42	46	56	58	61
Gender					
Male	186	187	191	186	187
Female	54	58	65	73	65
Race					
African	30	30	44	45	45
Indian	56	54	56	52	53
White	151	157	153	159	147
Academic position					
Professor	93	93	93	87	81
Senior lecturer	61	64	70	67	68
Lecturer	83	84	90	102	100

Parameter estimates for the zero inflation model were obtained using Stata. The strongly skewed nature of each of the plots that appear in Figure 1 seem to support the idea that the observed counts are being generated from a combination of two different sources – a non-publishing academic who has chosen to focus on the teaching aspect of their careers and will therefore never produce a research article throughout their entire academic career or a research active academic who in a so-called `dry year' may have recorded a zero record for that particular year.

Table 3 contains a set of parameter estimates for ϴ and γ. Any significantly positive value that occurs in γ can be associated with a factor that will serve to increase the log-odds ratio in favour of generating a structural zero (i.e. a person who will never publish a research article). Any significantly positive value obtained for ϴ should be interpreted as identification of a factor that will help to increase the expected per annum based research output of that individual when compared with a person from the baseline category (a white female who holds a senior lecturer position and has no PhD qualification).

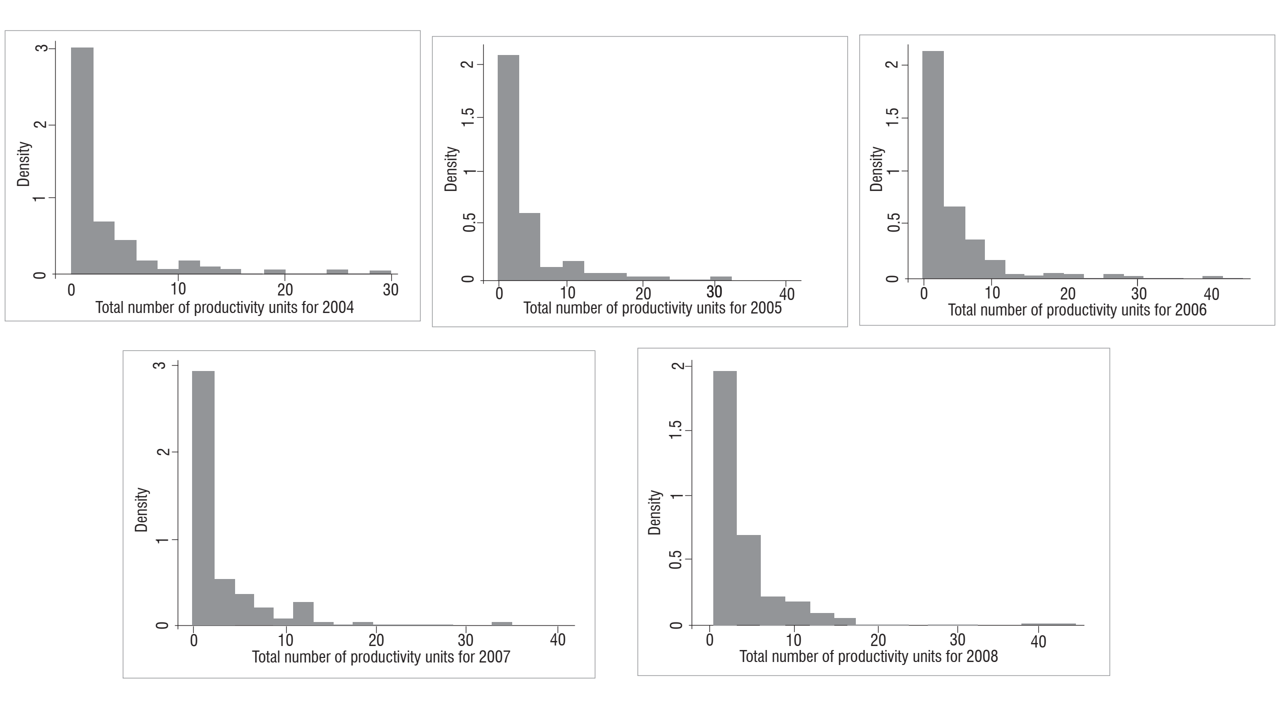


Figure 1: A set of observed sampling distributions for the number of productivity units generated in each year from 2004 to 2008

Table 3: Parameter estimates for Θ and γ in the zero-inflated Poisson model

	Parameter estimate Θ (SE)	Parameter estimate γ (SE)
Intercept term	1.12340(0.1540)*	0.2201(0.4599)
School size	0.0192(0.0015)*	-0.0138(0.0058)
Lecturer	-0.2878(0.0659)*	0.6630(0.1837)*
Professor	0.7627(0.0041)*	-0.9732(0.1706)*
Male	0.1015(0.0399)*	0.1078(0.1588)
African	-0.1148(0.0540)*	-0.0592(0.1913)
Indian	-0.2219(0.0475)*	0.5344(0.1628)*
Age	-0.1866(0.0211)*	0.1407(0.0832)
PhD	0.3934(0.0972)*	-0.8353(0.1917)*

**Denotes significance at the 5% level*

Focusing on those factors that help to distinguish someone who does publish from someone who does not, the estimates for γ that appear in Table 2 suggest that having a PhD, being a professor (rather than a lecturer) and being of non-Indian origin are the only variables that significantly increase the odds ratio associated with becoming an academic who does publish. More importantly, however, are the factors that do not seem to be affecting the odds ratio associated with becoming a research productive academic in the Faculty of Science and Agriculture at UKZN. In particular, age and school size do not seem to play a significant role in whether or not someone becomes a more research productive academic.

Focusing now on those academics who appear to be research active, the estimates for θ that appear in Table 3 suggest, at a 5% level of significance, that holding the position of professor, having a PhD qualification, being male and working in a large school all help to increase the research productivity of academics in the Faculty of Science and Agriculture at UKZN when compared with someone from the baseline category (a white female who holds a senior lecturer position in the faculty and has no PhD qualification). However, staff who are older seem to be less productive when compared with younger staff. Similarly, African and Indian researchers are not producing as much research output as their otherwise identical white counterparts.

Conclusion

The excess number of zero values observed in our data set may be attributable in part to a large number of academics within this data set who have chosen never to publish in their academic careers. Including these individuals in the analysis may impair the ability to focus on the problem of interest, which is to identify those factors that help to improve the publication rate of academics in the Faculty of Science and Agriculture at UKZN who have chosen to publish. The purpose of this study was to identify a specific set of factors that affect research productivity in the Faculty of Science and Agriculture at UKZN. Because the data set available does not explicitly identify someone as being a publishing or non-publishing academic, a method had to be developed to appropriately separate this pool of academics into those that publish and those that have elected to never publish. A set of covariates were then included in the model to establish their effect on research productivity.

In particular, having a PhD and working in a large school were found to have a significant impact on improving the research output of a publishing academic. If UKZN wants to become a research-focused university, non-publishing academics need therefore to be encouraged to complete a PhD degree.

References

1. Melck AP. Methods of financing universities with special reference to formula funding in South Africa [DCom thesis]. Stellenbosch: Stellenbosch University; 1982.

2. Venter RH. An Investigation of government financing of universities. 2nd ed. Report: SAPSE-110. Pretoria: Department of National Education; 1985.

3. Xie Y, Shauman K. Sex differences in research productivity: New evidence about an old puzzle. Am Sociol Rev. 1998;63(6):847–870. http://dx.doi.org/10.2307/2657505

4. Aksnes DW, Rorstad K, Piro F, Sivertsen G. Are female researchers less cited? A large-scale study of Norwegian scientists. J Am Soc Inform Sci Technol. 2011;62(4):628–636. http://dx.doi.org/10.1002/asi.21486

5. Kyvik S, Teigen M. Child care, research collaboration, and gender differences in scientific productivity. Sci Technol Hum Val. 1996;21(1):54–71. http://dx.doi.org/10.1177/016224399602100103

6. Barjak F. Research productivity in the Internet era. Scientometrics. 2006;68(3):343–360. http://dx.doi.org/10.1007/s11192-006-0116-y

7. Gonzalez-Brambila C, Veloso FM. The determinants of research output and impact: A study of Mexican researchers. Res Policy. 2007;36(7):1035–1051. http://dx.doi.org/10.1016/j.respol.2007.03.005

8. Kyvik S. Age and scientific productivity: Differences between fields of learning. Higher Education. 1990);19:37–55. http://dx.doi.org/10.1007/BF00142022

9. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992;34:1–14. http://dx.doi.org/10.2307/1269547

The Siloam Valley and Makuleni Village – the site of the hypothetical enhanced geothermal system discussed by Dhansay et al. in an article on South Africa's geothermal energy potential on page 88 (front cover). A researcher collecting water from a hot spring in the area (back cover). Photos: Taufeeq Dhansay.

APPLYING SCIENTIFIC THINKING IN THE SERVICE OF SOCIETY

Our vision is to be the apex organisation for science and scholarship in South Africa, internationally respected and connected, its membership simultaneously the aspiration of the country's most active scholars in all fields of scientific enquiry, and the collective resource for the professionally managed generation of evidence-based solutions to national problems.

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

T +27 12 349 6600/21/22 | **F** +27 86 576 9514

WWW.ASSAF.ORG.ZA