SOUTH AFRICAN
Journal of Science

Shale gas potential of the Main Karoo Basin

Externality costs of coal-based electricity

South African timber products in green building

Perspectives of wild medicine harvesters in Cape Town

DNA barcoding for pollen identification

volume 113
number 9/10

SEPTEMBER/OCTOBER 2017

eISSN: 1996-7489

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

SEPTEMBER/OCTOBER

eISSN: 1996-7489

SOUTH AFRICAN

Journal of Science

volume 113

number 9/10

Leader

When personal opinion trumps science
John Butler-Adam 1

News & Views

Ancient DNA comes of age, but still has some teenage problems
Alan G. Morris 2

The privileges and opportunities of a research sabbatical
Brenda Wingfield 4

Searching for a symbolic shipwreck in Table Bay: *Haarlem* (1647)
Bruno E.J.S. Werz, Willem H.B. Steenkamp & Mark R. Prowse 6

Book Review

Racism: The 'soft touch' does not work
Brenda Leibowitz 10

Human evolution and South African science: Darwin's hunch in context
Alan G. Morris 11

The complex origin of modern humans
Emily Holt 13

Commentary

The Extended BSc Programme: Performance of students in Chemistry
C. Robert Dennis, Rina Meintjes, Charlene Marais, Ricky Versteeg & Jannie C. Swarts 15

Towards an integrated ecological restoration approach for abandoned agricultural fields in renosterveld, South Africa
Sheunesu Ruwanza 19

Review Article

Pollination: Impact, role-players, interactions and study – A South African perspective
Annemarie Gous, Sandi Willows-Munro, Connal Eardley & Zacharias H. (Dirk) Swanevelder 23

Resilience processes in sexually abused adolescent girls: A scoping review of the literature
Sadiyya Haffejee & Linda Theron 31

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

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

Correspondence and enquiries
sajs@assaf.org.za

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

Cover caption
A female carder bee, *Pachyanthidium cordatum*, taking a rest on a twig before a foraging trip; she has no pollen among her yellow abdominal pollen carrying hairs (photo: Connal Eardley). Gous and colleagues review the importance of pollination and plant–pollinator interactions from a South African perspective in an article on page 23.

Research Article

Potential of interval partial least square regression in estimating leaf area index
Zolo Kiala, John Odindi & Onisimo Mutanga 40

The benefits of segmentation: Evidence from a South African bank and other studies
Douw G. Breed & Tanja Verster 49

The potential of South African timber products to reduce the environmental impact of buildings
Philip L. Crafford, Melanie Blumentritt & C. Brand Wessels 56

Perspectives of wild medicine harvesters from Cape Town, South Africa
Leif Petersen, Andrew M. Reid, Eugene J. Moll & Marc T. Hockings 64

Externality costs of the coal-fuel cycle: The case of Kusile Power Station
Nonophile P. Nkambule & James N. Blignaut 72

Exploring the relationship between entry requirements and throughput rates for honours students
Mike Murray 81

Deflating the shale gas potential of South Africa's Main Karoo basin
Michiel O. de Kock, Nicolas J. Beukes, Elijah O. Adeniyi, Doug Cole, Annette E. Götz, Claire Geel & Frantz-Gerard Ossa 87

When personal opinion trumps science

Hurricane Harvey (August/September 2017) has resulted in over 30 deaths; the loss of 100 000 homes and 1 million cars; and damage, private and public, that it is estimated will cost as much as USD190 billion to remedy. The President of the USA has (at the time of writing) committed just under USD10 billion to the State of Texas (not yet mentioning the devastation in eastern Louisiana). At the same time, he is proposing trillions of dollars in tax cuts for business and the wealthy, and is demanding that Congress approve the cost of a southern border wall, which, in effect (over the time of construction), is estimated by engineers to be USD200 billion, not to mention the cost of preventing people entering the USA by boat to circumvent the wall.

On 30 January 2015, President Barack Obama signed an Executive Order 'Establishing a federal flood risk management standard and a process for further soliciting and considering stakeholder input'. One of the results was the securing of reservoirs which, during the damage caused by Hurricane Harvey, held back flood waters so successfully that they backed up and spread waters westwards into homes, to the point that engineers were forced to evacuate residents below the reservoirs and open the floodgates.

It might have been logical to assume that the current President would have reflected on the benefits of the Obama Executive Order and so strengthened its requirements in view of the known probability of increasing risk to low-lying areas in the USA (home to almost 14 million people). In fact, 50% of the world's cities most at risk from coastal flooding are in the USA – Miami, New York, New Orleans, Tampa and Boston. Houston is not strictly a coastal city (although for 10–15 km it could very well be), but 7000 homes are located in areas that have been subject repeatedly to flooding over many years.

But the predicted risks were not part of the current President's understanding of the world and so, instead of strengthening Obama's Executive Order, he overrode it on 15 August 2017. On that day, he signed Executive Order 13807 – not strengthening Obama's Order but rolling back many of its conditions and requirements. Hurricane Harvey had already begun to form on 13 August 2017 and, 2 days after the signing, had reached the status of a tropical storm. In view of the course predicated for the storm, it is the kind of development that Presidents would be informed of in their daily briefings. The Executive Order had, however, been signed and the views of scientists are, in any event, held to be of little significance.

President Trump has since tweeted (on 30 August 2017) that 'After witnessing first hand the horror & devastation caused by Hurricane Harvey, my heart goes out even more so to the great people of Texas!'@realDonaldTrump.

But as the *Dallas Times* reported

> *With Houston still inundated, Trump kept away from the heart of the disaster zone. ...It wasn't just that he didn't visit a shelter or interact with any Texans who'd lost a home or loved one. He made no public mention of Texans undergoing such hardship.*

The problem is worse, of course, because his appointed head of the Environmental Protection Agency (EPA), Scott Pruitt, has previously been funded by large businesses in the fossil fuel industry. Although Pruitt says he does believe in global warming, he says equally clearly that he does not believe that there is a human cause, and that fossil fuels should be promoted in the interests of energy generation and economic growth. In his previous position as Attorney General of Oklahoma, Pruitt sued the EPA 14 times, leading Ken Kimmell, President of the Union of Concerned Scientists in the USA, to issue a statement that

> *...Pruitt is supported by the very industries he would be responsible for overseeing, particularly fossil fuel producers. He has questioned the clear scientific evidence for climate change and resisted science-based rules that protect our air and water from pollution. ...*

And it is now also clear that there is an intention to cut the EPA's staffing levels by over 30%, beginning with staff who work on climate change.

The effects of 'opinion vs science' sadly run deeper in society for a core of Americans, for while almost 60% of the population believes that there is such a phenomenon as global warming – and that it is mostly caused by human agency – a larger majority believes that the threats of climate change either will happen in the distant future, or will not happen to themselves or their families. This view is presumably not currently held by thousands of residents of Houston and other southeastern parts of the country.

The right-wing media in the USA are, for the most part, also sceptical of the human role in global warming and carry considerable public weight. Popular right-wing commentator Ann Coulter, for example, stated that she 'would be more willing to believe that gay people caused the flooding than buy into the idea of climate change' although quite how that might also apply to devastating flooding in South Asia, which has killed more than 1200 people in a swath from New Delhi to Dhaka, is not clear. It is equally unclear as to whether the broader, equally (or more) disastrous, flooding in Asia and elsewhere is the world has become part of the awareness of the President.

Simon Baptist, *The Economist*'s Global Chief Economist, a far more astute observer, had this to say:

> *With Hurricanes Irma and Harvey battering the Caribbean and southern US states, I have been thinking about the implications of climate change for countries across the world. The cost of relief and mitigation from natural disasters is rising, but the displacement of large communities in the longer term represents an international crisis in the making. The UN estimates that an average of 22.5m people a year have been displaced by natural disasters since 2008, yet this remains a trickle compared with the flood of refugees and migrants that could result from the growing effects of global warming. Large concentrations of people in India, Bangladesh and China are at risk from rising water levels and floods, while a loss of agrarian land, acidification of oceans and extreme weather events pose risks to communities everywhere.*

That such a clear-cut reality should be so muddied, at the expense of sound decision-making, is difficult to believe. That seems, however, to be the price to be paid for Presidential hubris and personal opinion.

HOW TO CITE: Butler-Adam J. When personal opinion trumps science. S Afr J Sci. 2017;113(9/10), Art. #a0235, 1 page. http://dx.doi.org/10.17159/sajs.2017/a0235

Ancient DNA comes of age, but still has some teenage problems

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

KEYWORDS:
South African archaeological sites; fossil DNA

HOW TO CITE:
Morris A. Ancient DNA comes of age, but still has some teenage problems. S Afr J Sci. 2017;113(9/10), Art. #a0232, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0232

In 2015, I wrote an opinion piece[1] in this journal announcing the successful extraction of a mitochondrial DNA (mtDNA) sequence from a 2300-year-old skeleton from Saldanha Bay, South Africa[2]. My concluding passage noted that the latest research was focused on nuclear autosomal DNA to provide data on admixture between populations and the impact of natural selection on specific genetic traits and that we could expect publication of articles on this topic in the near future.

Well, I was not far off. A few months later, a paper was published in *Nature* presenting the genome of a 4500-year-old individual from Mota Cave in Ethiopia.[3] The actual analysis was excellent, but there were problems in the comparative data[4] and one of the conclusions of the paper – that of a backflow of Eurasian ancestry affecting much of western and central Africa around three millennia ago – was subsequently rejected. The error was as a result of incompatible software in the analysis of genetic sequences from different comparative samples, which did not detract at all from the actual gene sequence retrieved. The Mota project was one of several happening on a range of ancient African DNA samples. A year earlier, Molebogeng Bodiba from the University of Pretoria was working at a lab in Zurich on South African samples from Early and Late Iron Age sites.[5] Unfortunately she has been unable to generate full mitochondrial sequences from her specimens, but she indeed has been the first to produce at least partial mtDNA data on Iron Age samples. As I mentioned in the 2015 paper, a small group of us at the University of Cape Town, in collaboration with the lab in Leipzig, has been working on samples from the Later Stone site of Faraoskop in the Western Cape, to find both mitochondrial and nuclear DNA samples and we are on the cusp of publication. Other Later Stone Age samples have been obtained from specimens in the Iziko Museums and are being analysed in Copenhagen by a different team; and most recently has been the report on three Later Stone Age foragers and four Iron Age individuals from Natal analysed in Uppsala, Sweden.[6]

Just to add to the mix are two studies from north of our border. Sloan Williams and Ryan Raaum[7] have extracted mtDNA sequences in a lab in Chicago from 38 individuals from the Swahili coast. The site of Mtwapa was a major Swahili town from the 9th to the 18th centuries, and the burials dating from the 17th century represent the largest sample of ancient DNA from a single site in Africa. This study is going to give us a wonderful insight into the biodynamics of a single population in the African past. Another team associated with David Reich's lab in Boston has extracted genomic DNA from several individuals from Tanzania, including an ancient skeleton from the islands of Zanzibar. And further to the west are two projects centred on Malawi and Zambia, looking at both Iron Age and Later Stone Age samples. The two studies are using different laboratories – one in Boston and the other in Uppsala. The Malawi project has successfully extracted full genomic DNA from a number of Later Stone Age specimens going back some 4000 years.[8]

This hive of African DNA analysis does not even include a wider range of research on the genetic lineages of living African peoples in search of the elusive point of divergence of our ancestors from the Neanderthals and other archaic populations and the subsequent divergence of populations in Africa.[9-14]

But there are problems. This rush to extract the secrets of ancient DNA in Africa has presented the curators of the collections of archaeological skeletons with ethical issues because the research requires the destruction of human bone. I have identified four central problems that concern me and that have been echoed in my private correspondence with various colleagues: competition between labs for samples; the danger of parachute research; the divorce between bioarchaeology and genetics; and laboratory methodologies and comparative data.

Competition between labs for samples

Competition for samples has become a very real problem. There are at least five labs that have been processing archaeological skeletons from South Africa. Back in May 2014, I made a list of all ancient DNA projects on South African specimens that had, up to that point, been proposed or were in action. I counted 13. Not all of these projects have taken place. In some cases permission to sample has been refused. The Department of Human Biology at the University of Cape Town, the Iziko Museums and the School of Anatomical Sciences at the University of the Witwatersrand all have elaborate procedures that must be followed before their respective osteological curators will grant permission to sample. In every case the science involved in the project (and the hypothesis it wishes to test) is required to meet an exacting standard. In the case of Iziko Museums there must also be a clear-cut benefit to descendant communities. One reason for refusal is that the project is simply an attempt to analyse skeletons because they are old and available. This may be good for the laboratory concerned, but it is just plain bad science and is perilously close to 'mining' of bone specimens from museums. In the case of very old specimens, such as Middle Stone Age material, the curating institutions have requested that associated animal bone is processed first and permission to analyse the human remains will only be granted if the animal specimens produce results. In addition, an export permit is required from the South African Heritage Resources Agency before any specimen is sent out of the country to a foreign laboratory.

The danger of 'parachute research'

Sampling in this kind of research is very easy to do. All that is required is a nubbin of bone and once the sample actually leaves the country, all of the analysis happens elsewhere. So how should South African researchers fit in? There has been an attempt to train young South Africans to work with ancient DNA and I am aware of students who have been sent to labs in Leipzig (Germany), Zurich (Switzerland) and Phoenix (Arizona, USA), but there are not as

yet any jobs for those who have qualified. For a number of years there was an active resistance to setting up a South African lab in the belief that it was too expensive and funding would be better spent on projects that have a more direct benefit to the previously disadvantaged people of the country. That attitude is now changing in some quarters and I have heard talk of setting up labs in Cape Town, Johannesburg and Pretoria. How would such labs link to overseas institutions?

The divorce between bioarchaeology and genetics

There has been a definite tendency for genetic research to ignore information from osteologically based previous research.[15] Until very recently, whether the data were Y-chromosome, mitochondrial or autosomal DNA, the search has been for lineage and the presence of an actual skeleton has been irrelevant. All that has been required is a tiny fragment of bone that can yield DNA, but can such studies give us a true picture of the past? The answer is 'yes' in terms of lineage, but 'no' in terms of life experience and adaptation.

This issue is indeed important because the first choice in sampling should be that from as complete a skeleton as possible so that we can compare genetic and osteological data. Perhaps the most extreme example of this problem is the construction of the human ancestor known as the 'Denisovans'.[16] Much has been written about the genetics of these distant ancestors, but all of it has been based on one finger bone and three teeth from one site. We actually know nothing about these people except for their genetic shadow. The forensic anthropologist in me screams that I must have a body before making any conclusions. The same goes for the discussion of people from the comparatively recent African past.

Laboratory methodologies and comparative data

As a non-geneticist, it did not cross my mind that different labs might produce different DNA results. Some years ago I had my own Y-chromosome and mtDNA analysed. The results were fascinating, but I was extremely surprised to discover that if I sent the same samples to different DNA heritage laboratories I could get different results. It is not the analysis itself that is different, but the reference samples that are chosen for comparison. Clearly this kind of problem can be resolved as the analysed samples become more numerous (as long as the different labs share their results), but I have recently discovered that not all labs are the same when it comes to long sequence autosomal DNA. The processing methods are not interchangeable and there are at least two different methodologies that produce different success rates and differing levels of data volume. How do we resolve this? One possibility suggested by an anthropological genetics colleague is to allow for multiple samples to be taken from the same individual. Basically we are saying double the destructive sample size. That would not be popular, but a new technique has just been published[17] that utilises a miniscule core boring of bone. Double sampling would not be such a tender point with such new methodology.

Where do we go from here? How do we resolve these problems? Much of the competition is publication driven with labs chasing the next *Nature* paper. I cannot overestimate the power of high-impact journal publications in generating funds for laboratories, or promotions for their denizens. But the result is that very little is ublished in African accessible journals. Where are the papers that engage the African scientists in the developing countries to the north of our border? The research results from the study of ancient DNA directly involve the descendant communities living in the countries from which the archaeological samples are extracted. And even when the populations being studied are extinct, the story of their presence is part of the heritage of all of the people who live there now.

Acknowledgement

Alan Morris is supported by an Incentive Grant from the National Research Foundation of South Africa.

References

1. Morris AG. Ancient DNA comes of age. S Afr J Sci. 2015;111(5/6), Art. #a0108, 2 pages. https://doi.org/10.17159/sajs.2015/a0108
2. Morris AG, Heinze A, Chan E, Smith AB, Hayes VM. First ancient mitochondrial human genome from a pre-pastoralist southern African. Genome Biol Evol. 2014;6(10):2647–2653. https://doi.org/10.1093/gbe/evu202
3. Llorente MG, Jones ER, Eriksson A, Siska V, Arthur KW, Arthur JW, et al. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. Science. 2015;350:820–822. https://doi.org/10.1126/science.aad2879
4. Callaway E. Error found in study of first ancient African genome. Nature News. 2016 January 29. https://doi.org/10.1038/nature.2016.19258
5. Bodiba M. Ancient DNA analysis of the Thulamela remains: Deciphering the migratory patterns of a southern African human population [MSc thesis]. Pretoria: University of Pretoria; 2014.
6. Schlebusch CM, Malmström H, Günther T, Sjödin P, Coutinho A, Edlund H, et al. Ancient genomes from southern Africa pushes modern human divergence beyond 1 260,000 years ago [bioRxiv preprint 145409]. c2017 [cited 2017 Aug 24]. http://dx.doi.org/10.1101/145409
7. Raaum R, Williams SR, Kusimba C, Monge J, Morris AG, Mohamed MM. Decoding the Swahili: The genetic ancestry of the Swahili. In: Wynne-Jones S, LaViolette A, editors. The Swahili world. Oxford: Taylor & Francis/Routledge. In press 2017.
8. Skoglund P, Thompson JC, Prendergast ME, Mittnik A, Sirak K, Hajdinjak M, et al. Reconstructing prehistoric African population structure. Cell. 2017;131, Art. #9774, 13 pages. http://dx.doi.org/10.1016/j.cell.2017.08.049
9. Barbieri C, Güldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, et al. Unraveling the complex maternal history of southern African Khoisan populations. Am J Phys Anthropol. 2014;153:435–448. https://doi.org/10.1002/ajpa.22441
10. Barbieri C, Hübner A, Macholdt E, Ni S, Lippold S, Schröder R, et al. Refining the Y chromosome phylogeny with southern African sequences. Hum Genet. 2016;135:541–553. https://doi.org/10.1007/s00439-016-1651-0
11. Chan EKF, Hardie R-A, Petersen DC, Beeson K, Bornman RMS, Smith AB, et al. Revised timeline and distribution of the earliest diverged human maternal lineages in southern Africa. PLoS ONE. 2015;10(3), e0121223, 17 pages https://doi.org/10.1371/journal.pone.0121223
12. Chimusa ER, Meintjies A, Tchanga M, Mulder N, Seioghe C, Soodyall H, et al. A genomic portrait of haplotype diversity and signatures of selection in indigenous southern African populations. PLoS Genet. 2015;11(3), e1005052, 28 pages. https://doi.org/10.1371/journal.pgen.1005052
13. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo R, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016;538:201–206. https://doi.org/10.1038/nature18964
14. Schlebusch CM, Prins F, Lombard M, Jakobsson M, Soodyall H. The disappearing San of southeastern Africa and their genetic affinities. Hum Genet. 2016;135:1365–1373. https://doi.org/10.1007/s00439-016-1729-8
15. Morris AG. Prehistory in blood and bone: An essay on the reconstruction of the past from genetics and morphology. Trans Roy Soc S Afr. 2005;60(2):111–114. https://doi.org/10.1080/00359190509520487
16. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010;468:1053–1060. https://doi.org/10.1038/nature09710
17. Sirak K, Novak M, Cheronet O. A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. BioTechniques. 2017;62(6):283–289. https://doi.org/10.2144/000114558

The privileges and opportunities of a research sabbatical

AUTHOR:
Brenda Wingfield[1]

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

CORRESPONDENCE TO:
Brenda Wingfield

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

KEYWORDS:
research leave; academia

HOW TO CITE:
Wingfield B. The privileges and opportunities of a research sabbatical. S Afr J Sci. 2017;113(9/10), Art. #a0234, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0234

As the title may suggest, I am currently undertaking a sabbatical – this time at Pennsylvania State University in the town of State College (Pennsylvania, USA). This sabbatical is the fifth that I have had the privilege to take during my 30-year academic career. The first was in 1994 at Iowa State University in Ames (Iowa, USA), the second at CSIRO in Canberra (Australia), the third at Murdoch University in Perth (Australia) and the fourth at the University of California, Davis in Davis (California, USA). They have all been very different and have without question added to my science, often in ways that I had not anticipated. I am reminded how important sabbaticals can be and take this opportunity to share some thoughts, particularly with my younger colleagues who might have doubts in this regard.

During my 7-year term in the dean's office at the University of Pretoria, I spent many hours encouraging academics, young and older, of the incredible benefits of taking a sabbatical 'offshore'. Some chose to take up this challenge with outstanding results; others have chosen to 'stay at home' and undertake research or writing locally. There are many reasons for the latter choice; some in my opinion are more valid than others. While the benefits of a sabbatical can be huge, leaving the comforts of one's home base is not for the faint hearted.

At the outset, allow me to explain some of what I achieved professionally during my various sabbaticals. One of the publications from my first sabbatical remains one of my most highly cited papers; I commonly use it as an example of how cutting-edge science can become routine in disease diagnostics – *but* the time lag for this outcome can be 10 if not 20 years, as it was in this particular case. My second sabbatical in Australia allowed me to produce one of the first papers on the phylogenetics of treating the plant pathogens known as 'rusts'.[1] That paper has not been one of my 'best sellers' but has maintained an average of two citations a year since its publication. It was an example of an idea that was far ahead of its time. It took 4 years before it was first cited but it probably will continue to be cited for many years to come.

Another short sabbatical in Australia allowed me the 'space' to think deeply about generic concepts for plant pathogenic fungi in *Ceratocystis*[2] and to begin the process towards developing a natural classification for this group. The genus has now been split into many different genera and we have robust DNA sequence data to support this new taxonomy. There are many other examples that I could share but the point is that research sabbaticals can have a very significant impact on one's scientific output and academic development. I have no doubt that I would not hold a SARChI Chair in Fungal Genomics without the knowledge that I gained from my various sabbaticals. And I am certain that my sabbaticals have also had a substantial impact on other things, such as my NRF rating.[3]

Many years ago, one of my close colleagues suggested that one should consider a sabbatical in the same light as one would when undertaking an extended camping trip. Most often one does not have to sleep in a tent but many 'creature comforts' are lost, albeit temporarily. Certainly my current sabbatical has highlighted this fact. I have a small furnished apartment on the bus route to the University. This means that I do not have access to a vehicle and it implies that shopping for groceries is more of a challenge than normal. I have learned rapidly to purchase only as much as one can carry comfortably. And, as during other sabbaticals, I have learned to make do with minimal home facilities. Minimalism is not only part of the challenge, but also brings freedom to focus on research and writing.

My sabbaticals have all been in countries with a stronger currency than South Africa – the country in which I choose to live – which has had some stressful consequences. I can remember that my children thought something terrible had happened to them when they were not able to enjoy their favourite foods. The value of the rand against the currency of the country you are in becomes frighteningly real. In the case of my present sabbatical, the rand value dropped by almost 20% to the dollar during my first 2 weeks in the USA. One can shop more frugally for food, but my accommodation became that much more expensive. This situation clearly puts many off taking sabbatical leave offshore, but these issues have been true for many years and there are usually ways to find the needed support to live, even if frugally.

My first sabbatical coincided with the first democratic elections in South Africa. At first, we heard a lot about the elections in the media, but then past president Richard Nixon died. From that point on, we had little news from South Africa – all news was about Nixon. My current sabbatical feels similar – many changes have occurred in South Africa since I left the continent. But these are different times. We have Internet connection and easy access to world news. I can also talk more easily with family, colleagues and students thanks to Skype, FaceTime and WhatsApp. Clearly, some of the challenges of sabbaticals of 10 and 20 years ago are no longer relevant. But one certainly still needs to accept a change of mindset; to me this mindset is also one of the driving forces of what I now term the 'academic sabbatical advantage'. One of the reasons that I have been such a strong proponent of research sabbaticals is that I believe in the value of being pushed out of one's comfort zone. One of the motivational quotes which resonates in this regard is: 'Outside of the comfort zone is where the magic happens.' Another saying is: 'The only difference between a rut and a grave is the depth of the hole.' Being outside one's comfort zone is important if one is to easily embrace new ideas and new ways of doing things (and potentially even magic).

The immediate value of being on sabbatical is that one immediately becomes almost 'anonymous' and has very few obvious responsibilities. Although this anonymity can feel strange, it – combined with a sudden lack of 'importance' – brings many advantages: time to think, to read, to write and to make new contacts. And perhaps one of the most important points of a sabbatical is to be in a position to take a step back and think about one's career, one's research and one's life – what one seeks to achieve in the short term and in the more distant future.

Acknowledgement

The National Research Foundation (South Africa) and the University of Pretoria are thanked for funding.

References

1. Wingfield BD, Ericson L, Szaro T, Burdor JJ. Phylogenetic patterns in the Uredinales. Australas Plant Pathol. 2004;33:327–335. http://dx.doi.org/10.1071/AP04020
2. De Beer ZW, Duong TA, Barnes I, Wingfield BD, Wingfield MJ. Redefining *Ceratocystis* and allied genera. Stud Mycol. 2014;79:187–219. http://dx.doi.org/10.1016/j.simyco.2014.10.001
3. Wingfield BD. The ABCs of an NRF rating. S Afr J Sci. 2014;110(7/8), Art. #a0072, 2 pages. https://dx.doi.org/10.1590/sajs.2014/a0072

Searching for a symbolic shipwreck in Table Bay: *Haarlem* (1647)

AUTHORS:
Bruno E.J.S. Werz[1,2]
Willem H.B. Steenkamp[1,3]
Mark R. Prowse[1,4]

AFFILIATIONS:
[1]African Institute for Marine & Underwater Research, Exploration and Education (AIMURE), Cape Town, South Africa
[2]Department of History and Heritage Studies, University of Pretoria, Pretoria, South Africa
[3]BroadBand Geophysical (Pty) Ltd, Shelley Point, South Africa
[4]Underwater Surveys (Pty) Ltd, Cape Town, South Africa

CORRESPONDENCE TO:
Bruno Werz

EMAIL:
ceo@aimure.org

KEYWORDS:
Dutch East India Company; geophysical survey; Dolphin Beach; South Africa

HOW TO CITE:
Werz BEJS, Steenkamp WHB, Prowse MR. Searching for a symbolic shipwreck in Table Bay: *Haarlem* (1647). S Afr J Sci. 2017;113(9/10), Art. #a0227, 4 pages. http://dx.doi.org/10.17159/sajs.2017/a0227

When on 25 March 1647 the VOC ship *Haarlem* stranded in Table Bay, nobody expected that this incident would become the catalyst that created one of the roots of current multiracial and multicultural South African society.[1] Of the ship's crew, 58 were repatriated soon after stranding, but 62 men stayed behind to try and salvage as much of the cargo as possible. During their sojourn, the men from *Haarlem* came into contact with indigenous people. Upon returning to the Netherlands, the crew reported favourably of their experiences. As a result, VOC management decided to establish a much-needed stopover for its ships that later developed into the City of Cape Town. Although no conclusive physical evidence of the wreck has yet been obtained, the multidisciplinary approach followed in an effort to locate the wreck of the *Haarlem* is reported here.

The basis is provided by historical information that is contained in archival documentation. Of particular importance are contemporary eyewitness accounts, as contained in part of a journal that was kept by the junior merchant from the *Haarlem*, Leendert Jansz, and associated correspondence.[2,3] Jansz was put in charge of the salvage attempts that followed the wrecking. Additional information could be abstracted from a report by commissioners who visited the wreck during the course of 1647; details provided by the commander of the fleet that repatriated the remainder of the crew in 1648; accounts by the first commander of the settlement at the Cape, Jan van Riebeeck; as well as a contemporary published description of the Cape of Good Hope.[4-8]

Wreck location

The documents provide information on the approximate location of the incident. A report, dated 29 August 1647 and compiled by officers from the ships *Tijger*, *Henriette Louise* and *Noord Munster* who visited during a stopover, indicates that: '…*Haerlem*…stranded at the north side of this bay'[4]. Another reference is provided by Jodocus Hondius III (1622–1655) who in 1652 stated: '… the bight that is called the bight of Sardanje, which is the east side of Table Bay. In the year 1647, the ship *Haarlem* beached in this bight…past and to the north of the Salt River'[8]. An accompanying map, showing the coastline from St Helena Bay to Cape Hangklip, indicates the 'Bogt van Sardanje' at the eastern side of Table Bay, to the south of the rock outcrops near present-day Blaauwberg, to the southeast of Robben Island and north of the Salt River.

Additional evidence as to the approximate place of foundering is provided by cartographic material that is lodged in the Nationaal Archief in The Hague. The first of these documents is a so-called fair sheet which contains a sketch of the Table Bay area, orientated to the west, together with soundings, indications of suitable anchoring grounds and coastal profiles of the surrounding mountains (Figure 1).[9] On this document, the approximate location where the wrecking took place is indicated by a description near the northeast coastline of Table Bay: 'Approximately at this place the ship *Haerlem* stayed'. As the curve of the eastern coastline on this document is exaggerated, an electronic copy was manipulated to obtain a best fit to coincide with the actual situation. This was accomplished by enlarging and distorting the fair sheet to overlay identifiable marks over the same marks on a modern South African Navy chart of the area.[10] The results of this manipulation indicate a position for the wreck close to or on shore, to the southeast of Robben Island and in the vicinity of present day Table View – Dolphin Beach.

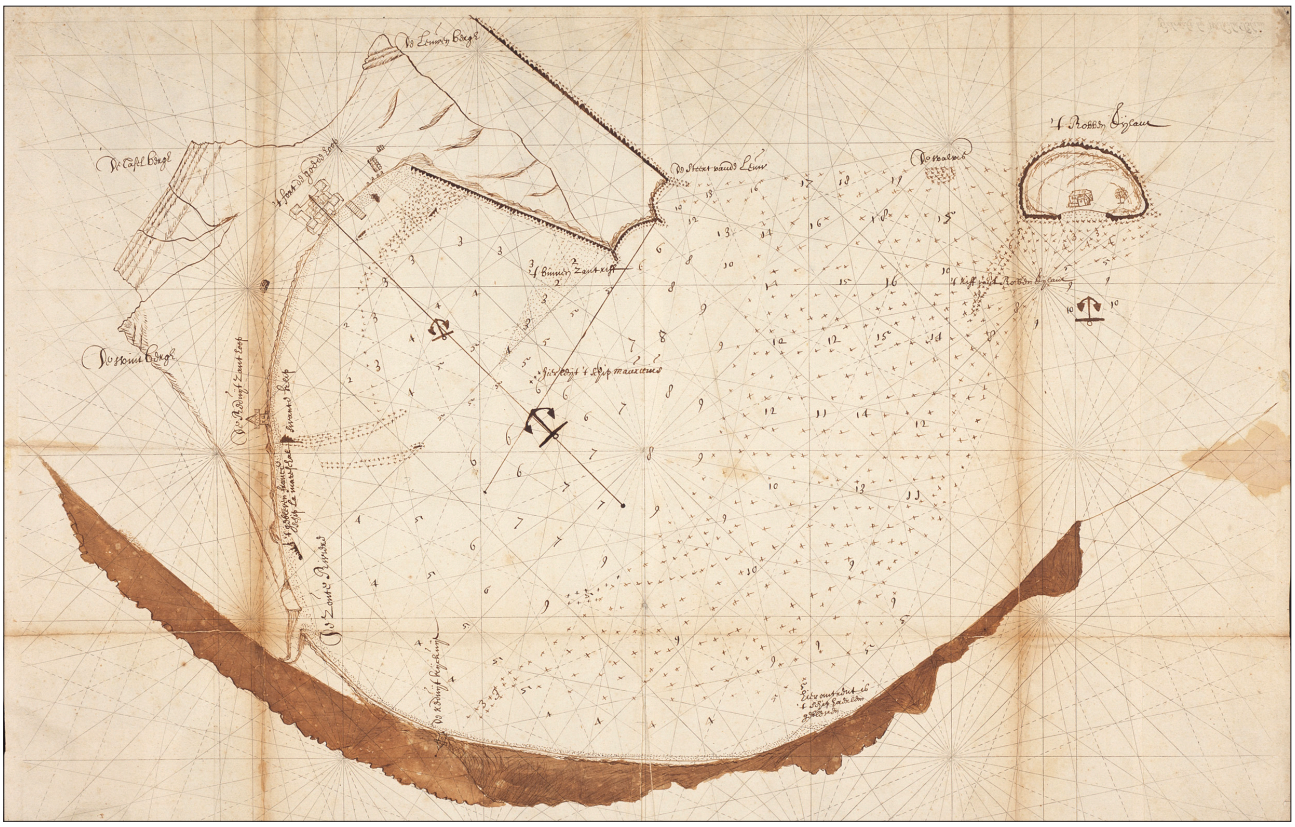
Two other contemporary images were traced that provide indications for the situation during the 1650s.[11,12] These manuscript maps indicate the location of the wreck in the northeastern section of the Table Bay area, as being either on[11] or very close[12] to the beach and opposite an area inland with salt pans. This position is confirmed by Van Riebeeck. In an entry in his journal, made on 20 May 1652, he states: 'Paid a visit to the wreck of the ship *Haerlem* which we saw still buried in the sand. …salt, which we found in fair abundance in that vicinity…' [6]. Nearly 7 years later, his journal reports that on 29 January 1659:

> *…at the mouth of the river which…our explorers called Hollands Rietbeecq* [Diep River]. *At its mouth this river is dry at present, but in the rainy season it flows strongly into the large salt-pans above the wreck of the Harlem* (sic) *at the tail of the Leopard Mountain, and then from there, through the pans, into the Salt River.*[7]

These salt pans are also mentioned in Hondius *Klare besgryving*:

> *Three or four miles up this Salt River there is a large patch of sand on which, in hot, dry weather sometimes so much pure white salt crystallizes (a hand's breadth deep), that it would be possible to take away a shipload of it.*[8]

The area that contained these salt pans is indicated on a map of Table Bay and the adjacent east coast by Joannes van Keulen II (1704–1755).[13] The map dates to 1753 and is very accurate, as was proven by overlaying it with the SA Navy chart 118 of Table Bay (Figure 2).[10] The overlay also made it possible to project the 17th and 18th northern and southern perimeters of the salt pans on the current situation. This shows that the northern boundary, when drawn along an east–west line, runs straight through the present Dolphin Beach Hotel in Table View. The southern line cuts across the northernmost houses in Milnerton. From this reconstruction it can also be concluded that the greater part of the old salt pans are currently covered by the Rietvlei Lake.



Nationaal Archief, The Hague, the Netherlands.

Figure 1: Fair sheet of Table Bay, dating to 1663, with the approximate position of the *Haarlem* wreck marked.

Photo: Mark Prowse ©AIMURE

Figure 2: Map of Table Bay and the adjacent coast, by Joannes van Keulen, dating to 1753, overlaid with SAN chart 118, Table Bay.

In the journal of Leendert Jansz, the place of foundering was recorded at a distance of: '1$^{1}/_{2}$ mile…from the Table Bay'[3]. It may be assumed that 'Table Bay' in this context refers to the old roadstead or the adjacent landing place, close to the location where the first official Dutch fort was constructed in 1652. At the time, 15 geographical Dutch miles equaled one degree of longitude on the equator. A standard of 7408 m for 1 Dutch mile was thus maintained for another reconstruction based on the length of one degree, which equals an average distance of 111.12 km. Taking the approximate location of Van Riebeeck's fort as a base, a distance of 11 112 m was plotted in a straight line across Table Bay. The end point of this line touched the coast at Dolphin Beach–Sunset Beach, directly opposite Rietvlei. Using the same starting point, another line with a distance of 11 112 m was projected following the old coastline as close as possible. This line ended slightly to the south of this point but still on Sunset Beach.

Geophysical surveys

A limited geophysical survey was undertaken during 2015, followed by a second phase that was undertaken during October–November 2016. During this last exercise, the full length of the permit area in the intertidal zone was covered, resulting in a total survey length of 25 949 m. The survey lines were approximately 10 m apart and the average magnetic sampling interval along the lines was approximately 1.2 m. Significant anomalies were scanned in greater detail by increasing the number of lines covering the areas of special interest. The results of this analysis, combined with additional information acquired since the start of the *Haarlem* project in 1989, indicate the following.

Two anomalies detected during the 2015 geophysical survey are caused by natural intrusive rocks – so-called dykes – containing magnetite. These dykes are clearly indicated by the countrywide airborne magnetic data set.[14] Geophysical modelling indicates the width of the dykes to be 10–20 m, with the tops of the dykes at depths of approximately -25 m. An anomaly found in between the dykes during the 2015 survey is probably caused by relatively modern refuse. During the 1970s, developments in Cape Town harbour necessitated sand winning. To that purpose, a pipeline was constructed from the harbour area along the beach to Rietvlei. It is thus quite possible that the anomaly is caused by pipes and other ferrous items that were not removed at the time.[15]

Other signals recorded in 2016 are caused by a power line or a pipeline in the northern section of the survey area or result from many additional anomalies, indicating that the sub-surface beach is littered with ferrous objects. Most of these anomalies have small footprints and geophysical modelling suggests that the ferrous objects are relatively shallow (<2 m) and probably weigh less than 20 kg. These are therefore judged to be low priority targets. However, there are five areas (A–E on Figure 3) where more significant anomalies with larger footprints occur. Geophysical modelling of these anomalies indicates the probable presence of ferrous objects of significant volumes that lie at depths of approximately 3–4 m. Figure 4 illustrates an example of the magnetic data that cover one of the principle target areas.

Although the wreck of the *Haarlem* has not been found to date, strong leads as to its approximate location have been obtained. Based on the available information, it can be concluded that the ship stranded in the northeastern section of Table Bay, specifically the Dolphin Beach–Sunset Beach area, opposite or very close to Rietvlei. The wreck must either be underneath the intertidal beach, or be in very shallow water at the foot of the beach. The geophysical surveys undertaken to date have indicated a number of anomalies. These anomalies are partly caused by geological features or are of an anthropogenic origin.

Photo: Billy Steenkamp ©AIMURE

Figure 3: Overview of the search area with five principal targets that need further investigation.

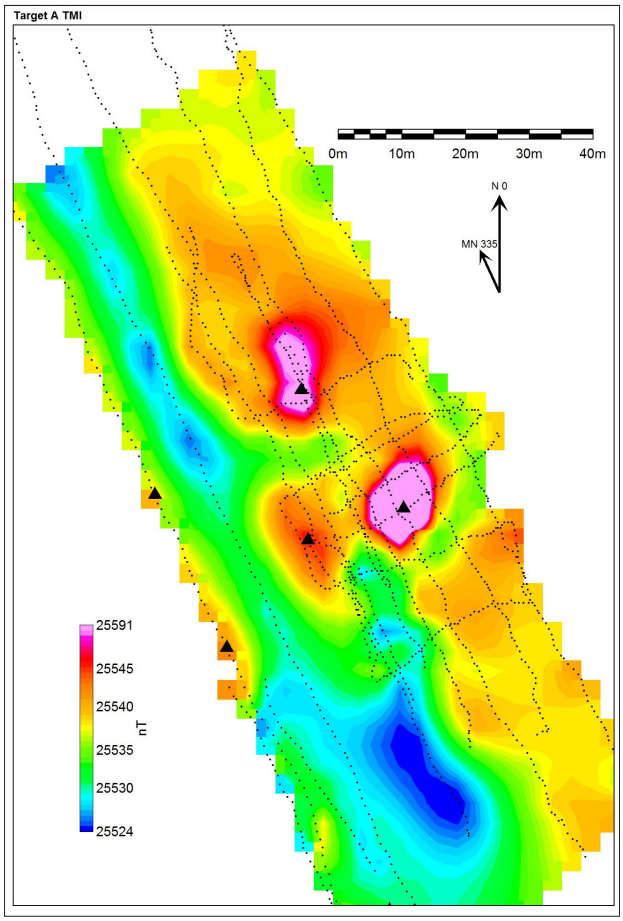


Photo: Billy Steenkamp ©AIMURE

Figure 4: Detailed magnetic survey of the principal target area, immediately to the west of the Dolphin Beach Hotel.

Five anomalous areas were recorded that indicate the probability of buried wrecks. One of these was found slightly north of the Dolphin Beach Hotel and this site was excavated by one of us (B.E.J.S.W.) in 1996. This wreck could not be identified at the time. Nevertheless, copper hull sheeting indicated its earliest period of construction as towards the end of the 18th century. Copper sheeting was only introduced from about the 1780s onwards, to prevent fouling of ship's hulls and to prevent deterioration caused by marine borers. The same applies to another wreck of which fragments were found after the American container vessel *Sealand Express* that stranded on Sunset Beach in August 2003 was pulled off. This is probably the same anomaly as indicated by Figure 3, point D.

An inventory of historical shipwrecks in the Blaauwberg, Table View, Rietvlei and Milnerton areas was compiled, using a variety of archival documents and other references.[16-20] The records indicate that at least 34 shipwrecks occurred in the area of interest. This number may possibly be higher, as at least 94 wrecks occurred in the Table Bay area for which the place of foundering is not specified.[20] Of the 34 wrecks whose approximate location is known, one dates to the 17th century (*Haarlem,* 1647). This ship is recorded as having foundered near the Table View–Sunset Beach–Milnerton area. Two shipwrecks dating to the 18th century (*La Cybelle,* 1756 and *Severe,* 1784), are reported to have been lost near Blaauwberg, further to the north. These ships would not have had copper sheeting on their hulls, with the possible exception of the *Severe*. The later wrecks – 31 of them – are spread over both areas and they would have either had hull sheeting or been constructed of iron. Once found, the *Haarlem* can be positively identified based on its location and because its hull was not fitted with copper sheeting. Furthermore, the wreck should still contain the remains of 19 iron cannons and four anchors, as indicated by the historical record.[5]

Acknowledgements

We thank the Board and Dive Unit of the African Institute for Marine and Underwater Research, Exploration & Education (AIMURE); Edwin Mills; BroadBand Geophysical; Ephan Potgieter and Andrew Matthews of Underwater Surveys; Allen Wood of the Koeberg Nuclear Power Station; Tristan Mostert of the Atlas of Mutual Heritage and Thomas Dresscher of the Nationaal Archief, the Netherlands; Professor Susan Webb and Dr Sharad Master of the School of Geosciences, University of the Witwatersrand; and Dr Pieter Kotze and Emmanuel Nahayo of the South African National Space Agency for their support and assistance. This survey was entirely sponsored by AIMURE, BroadBand Geophysical (Pty) Ltd and Underwater Surveys (Pty) Ltd.

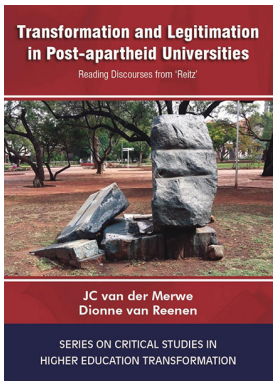
References

1. Werz B. The *Haarlem* shipwreck (1647): The origins of Cape Town. Pretoria: Unisa Press; 2017.
2. Nationaal Archief, The Hague, Archive of the United East India Company (VOC) 1166. Overgekomen brieven en papieren uit Indië aan de Heren XVII en de kamer Amsterdam, 1648, tweede boek, vervolg [Letters and documents sent from the Dutch East Indies to the VOC directors and the regional office of Amsterdam, 1648, second volume, continuation]. Dutch.
3. Western Cape Archives. Verbatim copies (VC) 284, Journal and letters of Leendert Jansz after the wreck of the ship 'Nieuw Haarlem' at the Cape, 1647.
4. Western Cape Archives. VC 284, nr. 6. Rapport, bijde Gecommitteerde vande gem[elde] scheepen gedaen, over haere gedaene visite aen Cabo de Bona Esperance, bij die vant [ver]ongeluckte schip Haerlem [Report by the commissioners of the mentioned ships, regarding their visit at the Cape of Good Hope to those of the wrecked ship Haerlem]. Dutch.
5. Nationaal Archief, The Hague. VOC Collection Wollebrant Geleynse de Jong 1.10.30.
6. Van Riebeeck J. Journal of Jan van Riebeeck I. 1651–1655. Thom HB, editor. Cape Town/Amsterdam: A.A. Balkema; 1952.
7. Van Riebeeck J. Journal of Jan van Riebeeck III. 1659–1662. Thom HB, editor. Cape Town/Amsterdam: A.A. Balkema; 1958.
8. Hondius J. Klare besgryving van Cabo de Bona Esperanca, 1652 [Clear description of the Cape of Good Hope, 1652]. Dutch.
9. Nationaal Archief, The Hague, Collection Leupe nr. 178. Kaart van de Tafelbaai, het Robben Eiland, de Klip de Walvisch, benevens aanwijzing waar eenige schepen gebleven zijn, 1663 [Chart of Table Bay, Robben Island, Whale Rock, as well as indications where some ships foundered, 1663]. Dutch.
10. South African Navy. Chart 118. Table Bay. 1:150.000 National series.
11. NA VEL 0803. In: Atlas of Mutual Heritage [database on the Internet]. Available from: http://www.atlasofmutualheritage.nl/en/
12. NA VEL 0804. In: Atlas of Mutual Heritage [database on the Internet]. Available from: http://www.atlasofmutualheritage.nl/en/
13. Van Keulen J. Kaart van de Tafelbaay vertoonende de reede van C. de Goede Hoop 1753 [Map of Table Bay showing the roadstead of the Cape of Good Hope 1753]. Dutch.
14. Council for Geoscience. Open file 1 km line-spacing dataset.
15. South African Railways and Harbours. Table Bay harbour. Harbour extension stage 2. Rietvlei borrow pit general layout, sheets 1–4, 29-7-1971. Cape Town: South African Railways and Harbours; 1971.
16. South African Library. List of South African shipwrecks. 1990. Unpublished.
17. Marine casualties southern African waters 1552 to 1913. S Afr Ship News Fish Ind Rev. 1982;37(6):18–42.
18. Marine casualties southern African waters 1914 to 1945. S Afr Ship News Fish Ind Rev. 1983;38(6):3–13.
19. Marine casualties southern African waters 1946 to 1984. S Afr Ship News Fish Ind Rev. 1984;39(6):3–27.
20. Werz BEJS. Strategic environmental assessment (SEA) for the Port of Cape Town and environmental impact assessment (EIA) for the expansion of the container terminal stacking area. Specialist study in maritime archaeology. Stellenbosch: CSIR Environmentek; 2003.

Racism: The 'soft touch' does not work

BOOK TITLE:
Transformation and legitimation in post-apartheid universities: Reading discourses from 'Reitz'

BOOK COVER:



AUTHORS:
J.C. van der Merwe and Dionne van Reenen

ISBN:
9781920382605 (softcover)

PUBLISHER:
SUN PRESS, Bloemfontein; ZAR350

PUBLISHED:
2016

REVIEWER:
Brenda Leibowitz

EMAIL:
Brendal@uj.ac.za

AFFILIATION:
Teaching and Learning, Faculty of Education, University of Johannesburg, Johannesburg, South Africa

HOW TO CITE:
Leibowitz B. Racism: The 'soft touch' does not work. S Afr J Sci. 2017;113(9/10), Art. #a0230, 1 page. http://dx.doi.org/10.17159/sajs.2017/a0230

The authors of this book – J.C. van der Merwe and Dionne van Reenen – both work for the Institute for Reconciliation and Social Justice at the University of the Free State, in Bloemfontein, South Africa. It was at this university that the infamous 'Reitz incident' occurred in 2008, in which a group of university students enlisted the participation of a group of black, mainly female, workers, in a video which mocked the workers without them realising it immediately, and which mocked the idea of racial integration. One would assume that the authors of this book would be highly invested in the prevention of such crudely racist acts at this university, and in the country – the book should contribute towards an understanding of how to create the kind of milieu in which racism does not occur. So does this book succeed in this task?

If one believes that careful reporting, sober analysis based on ethics, as well as theoretical readings are necessary to understand racist behaviour and to contribute to its prevention, then this book certainly succeeds. The book is divided into six chapters which cover a historical period leading up to the production of the racist video, the screening of the video and the aftermath. It carefully documents the events, using interviews with eight staff and four student leaders, minutes from internal meetings, and reporting from the media. In order to analyse various dimensions of the events, it draws from several theories: critical theory discourse and the work of Foucault and Hook; legitimation theory and the work of Habermas; and postural theory, drawing from the work of Johann Visagie.

From the detailed reporting and analysis in the book, two impressions remain. Firstly, that the Reitz incident was not an isolated incident. It was typical of much sexist and reactionary – but according to the authors, primarily racist – behaviour on the campus. This culture was tolerated, encouraged and even enforced by the university practices, for example by 'hazing' of first-year students in the residences. Secondly, it is very evident from the book that the forces of racism and intolerance did not emanate solely from the students, but also from amongst academics, parents and outside political formations. It was in response to these endemic trends that the leadership of the university at the time of the production of the video could have been seen as too soft.

I recognised many of the debates and tensions described in the book from an institution at which I worked for 10 years – although it was not as crudely racist as UFS. The authors seem to have taken care not to be sensational or overly essentialising and judgemental (the reporting is possibly even underwhelming in places). They bring many of the complexities and nuances of the situation out carefully.

The authors advance the idea of a legitimation crisis as a reason for the crisis, in that there was not a firm enough basis with which to establish dialogue within a robust democratic culture:

> *It is suggested here that mature, differentiated, discursive actions required for legitimation were not present in the UFS on a large enough scale because there was extreme hesitancy in adopting and applying democratic values. For this reason, from the first mumblings of 'transformation' on campus, the UFS was rendered vulnerable and open to the threat of a legitimation crisis and it is still not free from that hazard.* (p. 152)

The authors propose dealing with the lack of legitimacy by the following means: changing the institutional culture; instituting a rights-based approach; creating space for 'being political' on campus; doing anti-racism work; and establishing pre-conditions (p. 247). The elaboration of the final action – establishing pre-conditions – is the most interesting. It includes:

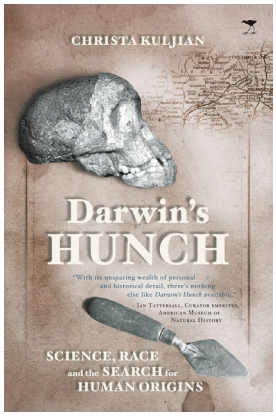
1. An effective governing apparatus through which collectively binding decisions can be implemented.
2. A clearly defined 'self' for the purposes of self-determination and self-transformation to which collectively binding decisions can be ascribed.
3. There must be a membership that can be mobilised for participation in institutional opinion-formation and will-formation orientated to the common good.
4. There must be an educational and social milieu in which a democratically programmed administration can provide legitimacy-enhancing steering and organisation. (p. 265)

These four pointers suggest what legitimacy of leadership, on a sound democratic basis, could look like. This legitimacy did not exist in the period leading up to and at the time of the Reitz incident, partly because of the institutional culture, partly because of the lack of a critical mass of progressive students and staff that could be mobilised, and partly because leaders did not see the need for a more robust and less evolutionary process to unleash the sorely needed process of transformation at the university. A firm, robust and visionary approach towards transformation might be needed at the UFS, but a similar point can be made for the other previously white, Christian, Afrikaans universities in South Africa. Surely the need for a less-than-soft touch in a democratic context with a clearly defined 'sense of self', could be offered as a solution at many other South African universities, including the previously white, liberal English universities? One could argue that in the light of the 2015–2016 student protests against fees and colonial education, there has been a crisis of legitimation at all South African public higher education institutions, and that this less-than-soft touch in a democratic context, is required in the system as a whole.

Human evolution and South African science: Darwin's hunch in context

BOOK TITLE:
Darwin's hunch: Science, race, and the search for human origins

BOOK COVER:



AUTHOR:
Christa Kuljian

ISBN:
9781431424252 (softcover)

PUBLISHER:
Jacana Media, Johannesburg; ZAR295

PUBLISHED:
2016

REVIEWER:
Alan G. Morris

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

EMAIL:
alan.morris@uct.ac.za

HOW TO CITE:
Morris A. Human evolution and South African science: Darwin's hunch in context. S Afr J Sci. 2017;113(9/10), Art. #a0231, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0231

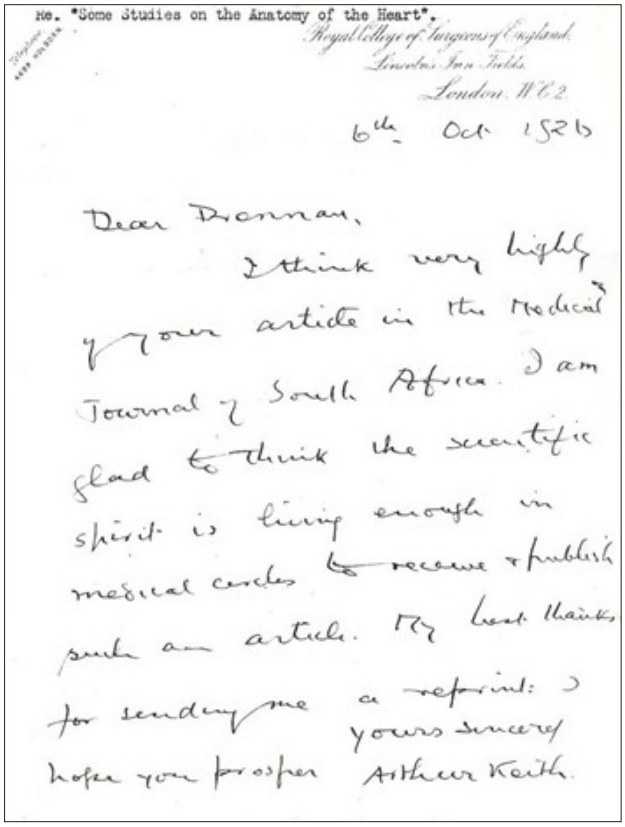
Back in 2015 I had the opportunity to spend 5 months in the UK searching for skeletons of Africans who had been accessioned in the anthropology collections of English, Irish and Scottish museums and universities. This was a project that I had been working on sporadically for many years and, thanks to the Leverhulme Trust, I was funded to be a visiting professor at Cambridge University and I made use of Cambridge as a UK base in order to do a final accounting if I could. One of the highlights of the research end of the trip was sitting in the library of the Royal College of Surgeons in London going through the pre-World War II letter files. This era was the heyday of skeleton collecting for the College and was only really terminated by German bombs hitting the college during the War. Although much was salvaged, the College lost its interest in anthropology and transferred the bones that survived the destruction to the Natural History Museum.

The correspondence through which I was paging included letters and notes that either referred to the arrival of the skeletons or described (sometimes in quite stark detail) how they were 'gathered'. My account of this correspondence is going to have to wait until I have had time to go through the huge mass of information I gathered, but what I wanted to focus on here was the South African connection that kept appearing in the correspondence files. From the English side, the main correspondent was Sir Arthur Keith, then the pre-eminent anthropological anatomist in the UK. It is hard to explain exactly how important this man was in his time, but basically he was the British Empire authority on human evolution. His support meant your success in publication and academic promotion, so it was intriguing to me to find South African names in his list of correspondents. Some were 'armchair' anthropologists who sent him boxes of bones from one source or another. Others were budding scientists who were based in what was then the still very new universities of South Africa. Keith was well acquainted with Raymond Dart, for he had written (grudgingly) in support of Dart's appointment in Johannesburg in 1923, but other names kept cropping up in the files. Chief amongst these was Matthew Drennan, my academic ancestor at the University of Cape Town.

This was not the first correspondence I had seen between Keith and Drennan. I had come across other letters between them in the old files of the University of Cape Town's Department of Human Biology. Although there is not a lot of scientific importance in their correspondence, the letters provide a real window into how a young academic in a distant realm of the far-flung empire saw himself in relation to the great authority of the day. Drennan had arrived in Cape Town in 1913 as a lecturer in anatomy, fresh off his medical training at the University of Edinburgh. It took quite a while for him to find his feet in Cape Town and research was not his first priority. He had no previous interest in anthropology, but he was certainly aware of the state of the science given the long history of anthropological research in Edinburgh's Department of Anatomy. By the 1920s, Drennan had begun his research career in studying the skeletonised remains of South Africans and their prehistoric predecessors and much of his correspondence with Keith was about the papers that he had begun to write on the subject. One letter, dated 3 May 1929, implores Keith to recommend publication of his new report on the Cape Flats skull in the *Journal of the Royal Anthropological Institute*. Drennan writes: 'We could give it to the Royal Society here, but then there would be no one to criticise it and few to read it.'

In many ways, this sentence sums up what was happening in the field of physical anthropology in South Africa in the years before World War II. These years were heady days of discovery, but what was critical was that these discoveries needed to be brought to the attention of the world. There were two key research thrusts: one was on the origin and variation of aboriginal South Africans, especially the Khoesan; while the other was on fossil discoveries that took the whole of humanity back to its roots. The 1920s and 1930s were indeed formative years, but what is important to understand is how the anthropological research was influenced by the colonial attitudes of the day and how these sometimes very racist ideas shaped the anthropological research of today. Saul Dubow[1] has looked at many of these issues, as have I[2], but what has been missing is someone to track through the more modern evidence and how our earlier social and academic biases and interests have impacted on current research. Now this has been done and what a superlative job it is!

Darwin's Hunch by Christa Kuljian is an important book because the search for fossil humans and the understanding of human variation is so important to South Africa. As the title of Kuljian's book indicates, it was Charles Darwin who first expressed the belief that the place where humans first evolved was Africa, but her book is not about Darwin. She instead focuses on how the question of race was mixed with the issue of evolution right from the start. Dubow[1] has touched on this issue, but what he did not do was to look in any depth at the personality of the players involved. This is Kuljian's big strength. She focuses on the Johannesburg researchers, partly because that is where she is based, but also because the University of the Witwatersrand has played, and continues to play, such a huge role in the search for the earliest humans. The first third of the book covers the period I mention above with a special focus on Raymond Dart. Two parallel themes emerge which continue in one form or another throughout the book. One theme is about the discovery of the fossil evidence – in this case the skull of the Taung child in 1925. The second is about the corresponding study of African people which provided a framework for the interpretation of the fossils. In particular was the dissection of /Keri/Keri, a young woman from the southern Kalahari whom Dart had met in life, and in death had obtained her body to study as a type specimen. She was the embodiment of what he considered to be the 'living fossils' of human evolution – the San of southern Africa. This part of the book ends with the unveiling of Piltdown as a forgery and the removal of this English fossil (championed by none other than Sir Arthur Keith) from the record of human evolution. The demise of Piltdown brought the African evidence to the fore and set the stage for the pre-eminence of Africa in the story of human origins.


Re. "Some Studies on the Anatomy of the Heart".
Royal College of Surgeons of England
Lincoln's Inn Fields
London. W.C.2.
6th Oct 1926
Dear Drennan,
I think very highly your article in the Medical Journal of South Africa. I am glad to think the scientific spirit is living enough in medical circles to receive & publish such an article. My best thanks for sending me a reprint: Yours sincerely Hope you prosper Arthur Keith.

Personal letter from Arthur Keith to Matthew Drennan, 6 October 1926.

Kuljian notes that the end of the Piltdown coincided with the rise of a new generation of African scholars, in particular Phillip Tobias. Tobias, who died in 2012, has left an outsized mark in the South African academic scene, but Kuljian sees him as an enigma. She identifies a paradox between his opposition to apartheid and his scientific practices which in many ways echo those of his mentor Raymond Dart. This is an important theme of the book. Kuljian has tried to go beyond the simple recounting of investigative tales. In particular she has teased out the stories of the 'small' people of science – the field workers, the lab assistants and, in particular, the women researchers. Her view ensures that we know they are not 'small' because of lack of importance, but because they have been long overlooked. Especially important to her is the fate of /Keri/Keri's skeleton which is not only a reminder of how poorly the earlier anthropologists treated the native populations, but is also a continuing saga because the remains themselves have gone missing.

The last few chapters of the book bring us up to events of the last few years, such as the discovery, excavation and unveiling of *Homo naledi* and the rise of ancient DNA. What Kuljian has done for those of us still working in the field, is to make us ask awkward questions of ourselves. Back in the days when Drennan was writing to Keith, the questions were about how South African scientists could make a mark on the world stage, but we have long been able to demonstrate our academic expertise as scientific equals to the academics of London and elsewhere. Kuljian now forces us to examine our science to see if it is not only correct, but appropriate and socially just. We know how we are received in the scientific journals of the world, but what does our research mean to South Africans themselves, especially those who have been excluded in the past? As Kuljian herself says, the old days of racial science are gone, but we still need to learn that science can never be divorced from the world around us, and that scientists are as much affected by the world in which they live as that world is affected by us.

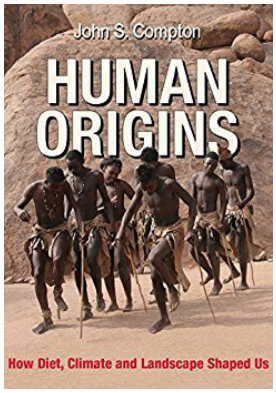
References

1. Dubow S. Illicit union: Scientific racism in modern South Africa. Cambridge: Cambridge University Press; 1995.
2. Morris AG. Biological anthropology at the southern tip of Africa: Carrying European baggage in an African context. Curr Anthropol. 2012;53(S5):S152–S160. https://doi.org/10.1086/662289

The complex origin of modern humans

BOOK TITLE:
Human origins: How diet, climate and landscape shaped us

BOOK COVER:



AUTHOR:
John S. Compton

ISBN:
9780620696036 (ebook)

PUBLISHER:
Earthspun Books, Cape Town; ZAR75

PUBLISHED:
2016

REVIEWER:
Emily Holt

AFFILIATION:
Previous: College of Science and Engineering, James Cook University, Cairns, Australia

Current: Department of Infectious Disease, Alfred Health, Melbourne, Australia

EMAIL:
emily.holt@my.jcu.edu.au

HOW TO CITE:
Holt E. The complex origin of modern humans. S Afr J Sci. 2017;113(9/10), Art. #a0233, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0233

Human origins and the factors influencing our evolution are fundamental to understanding our past, reflecting on our present, and determining our future. The seminal work of Charles Darwin provided evidence for a natural process of evolution, and an explanation for the immense abundance of species on earth; it also opened the way for study on the origin of humans.

In his book *Human Origins: How Diet, Climate and Landscape Shaped Us*, Compton includes a vast amount of information in support of the influence of the environment on the formation of life and evolutionary processes, as well as the associated physical traits, diet and behaviour of species. Tied to these are detailed discussions on DNA and genetics, hominin evolution and the most likely area in Africa that was the birthplace of modern humans (*Homo sapiens*). Compton also delivers a sobering reminder of the effects modern humans have had, and continue to have, on the environment, and where we may be in the future.

This book is comprehensive and the content is reflective of the author's more than 20 years in academia, knowledge of geology and location within South Africa. Compton covers a large range of subject matters in an easily understandable manner, and the arguments, explanations and evidence are well presented. The addition of figures, including photographs of fossils, further assist the reader by providing appropriate visual examples of explanations in the text. However, there are a number of areas where expansion of the subject matter would have provided clarity, and ensured that the reader is aware of the different interpretations of anthropological, archaeological and palaeoclimate studies. *Human Origins* is a great starting point for understanding evolution and the complexity of climate–environment–animal–human interactions; it should also entice the reader to expand their knowledge on the topics covered in the book.

Starting at the very beginning, Compton guides us from the Big Bang to the evolution of life on earth. The first two chapters are engaging, with a description of earth's formation, early geological and atmospheric processes, the appearance of life, and the perilous journey of evolution. Devoting these first chapters to pre-human origins is effective, as this background information is fundamental to understanding the evolutionary processes discussed later in the text.

The majority of *Human Origins* is dedicated to the journey from our earliest ancestors to *Homo sapiens* as the dominant species living on earth today. Compton delves into the myriad factors that are posited to have influenced the development of modern humans, including those stipulated in the title (diet, climate and landscape), as well as social and behavioural aspects. There is a particular focus on Africa, and for good reason: Africa has fossils of multiple hominin species, in addition to archaeological artefacts that demonstrate the adoption and progression of fire and tool use, as well as cultural development. Furthermore, Africa is, by consensus and supported by genetic studies, considered to be the birthplace of *H. sapiens*.

It is difficult to amalgamate large amounts of scientific information and Compton must be commended for doing so. However, there are areas in *Human Origins* that would benefit from the addition of more information; here follow some examples. Dating is vital for matching climate, environment and fossils, and placing them at key points in time; yet, there is a curious lack of discussion about the methods used to provide the dates presented. Also noted is the brief coverage of East Africa in the context of the discussion on isolation hypothesis/allopatric speciation, which is proposed as a major driver in *H. sapiens* emergence. While it is certainly true that the complexity of climate and environmental variation in Central and East Africa during glacial/interglacial periods makes the interpretation of isolated areas here difficult (which Compton acknowledges), the important hominin fossil sites in East Africa warrant further examination. Finally, when discussing modern human expansion, Compton mentions that Aboriginal Australians, soon after their arrival in Australia, played an exclusive role in altering vegetation type and causing megafauna extinction. I submit that ongoing shifts in climate, in addition to natural fires, have not been discounted as significant influences on the Australian environment and biota at that time.

As a final point, evidence and scientific consensus on modern human origin and ancestral lineage is in constant flux. With each new fossil discovery and improvement in dating methods, the proposed phylogenetic tree changes; Compton himself makes the comment that interpretations are likely to change markedly over time. It is unfortunate that progress is somewhat hindered by a lack of adequately preserved fossils from key areas in Africa. Nevertheless, recently reported studies have included some momentous discoveries. Fossils from Jebel Irhoud in northern Africa are now reported to be the oldest known fossils of what is identified to be early *H. sapiens,* with a stated age of over 300 000 years.[1,2] Additionally, dating of the *Homo naledi* fossils from Rising Star Cave in South Africa indicates that these fossils are much younger than first thought, with an age range of ~230 000 to ~330 000 years.[3,4] These findings will undoubtedly lead to further discussion and scholarly debate on hominin interactions, lineage and movement within, and out of, Africa.

I would certainly recommend *Human Origins* as a valuable text to readers with an interest in, or studying, the origins of life, evolutionary processes, hominin phylogeny, and the history and impacts of *H. sapiens*. Compton has written a book with easily understandable content, which introduces the reader to vital concepts and hypotheses in evolution, and the importance of environmental factors on human origins. The reader should also take note of the suggested readings at the end of the book – there is a list of material to challenge even the most intrepid scholar – to continue their journey to understanding the many factors leading to us as the only remaining species of the *Homo* lineage.

References

1. Hublin J-J, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. Nature. 2017;546:289–292. http://dx.doi.org/10.1038/nature22336
2. Richter D, Grün R, Joannes-Boyau R, Steele TE, Amani F, Rué M, et al. The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. Nature. 2017;546:293–296. http://dx.doi.org/10.1038/nature22335
3. Dirks PHGM, Roberts EM, Hilbert-Wolf H, Kramers JD, Hawks J, Dosseto A, et al. The age of *Homo naledi* and associated sediments in the Rising Star Cave, South Africa. eLife. 2017;6, e24231, 59 pages. https://doi.org/10.7554/eLife.24231
4. Hawks J, Elliott M, Schmid P, Churchill SE, Ruiter DJd, Roberts EM, et al. New fossil remains of *Homo naledi* from the Lesedi Chamber, South Africa. eLife. 2017;6, e24232, 63 pages. https://doi.org/10.7554/eLife.24232

The Extended BSc Programme: Performance of students in Chemistry

AUTHORS:
C. Robert Dennis[1]
Rina Meintjes[1]
Charlene Marais[1]
Ricky Versteeg[1]
Jannie C. Swarts[1]

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

CORRESPONDENCE TO:
C. Robert Dennis

EMAIL:
crobbie.dennis@gmail.com

KEYWORDS:
school performance profiles; graduates

HOW TO CITE:
Dennis CR, Meintjes R, Marais C, Versteeg R, Swarts JC. The Extended BSc Programme: Performance of students in Chemistry. S Afr J Sci. 2017;113(9/10), Art. #a0229, 4 pages. http://dx.doi.org/10.17159/sajs.2017/a0229

Admission to study for a BSc degree at South African universities is determined by a student's performance in the final Grade-12 National Senior Certificate (NSC) examination. These entrance requirements are:

- a specific admission point (AP score), calculated from the subject achievement levels obtained in the final Grade-12 NSC examination;
- a degree endorsement; and
- specific achievement or performance levels in Mathematics and Physical Sciences.

The current entrance requirements[1] for a BSc degree at the University of the Free State (UFS) are:

- a NSC Grade-12 certificate with a degree endorsement;
- a minimum AP score of 30;
- an achievement Level 4 in the Grade-12 language of instruction;
- Grade-12 Mathematics at a minimum achievement Level 5; and
- Grade-12 Physical Sciences at a minimum achievement Level 4 and/or Grade-12 Life Sciences at a minimum achievement Level 5.

The requirement of Physical Sciences and/or Life Sciences is dependent on specific study programmes.

To cater for school-leavers who do not meet these entrance requirements[1], the UFS implemented the Extended BSc Programme (for those with AP scores of 25 to 29) in 2005 to complement the very successful science stream of the University Preparation Programme (UPP)[2-4] (for those with AP scores of 20 to 24). In 2011, the Extended BSc Programme was moved from the Bloemfontein Campus of the UFS to the South Campus where the UPP has been situated since 2009.

In the first year of study in the Extended BSc Programme, students initially had to register for 32 mainstream credits in Chemistry and 32 foundation Mathematics credits, similarly to the UPP.[2,4] Students also had to register for development modules in Academic Literacy, Concepts in General Science and Introduction to Computer Usage. In 2011, the module 'Concepts in General Science' was replaced by the module 'Skills and Competencies for Life Long Learning: Natural Science'. Mainstream credits in Biology have been added to the first year of the programme since 2014 to bring the Extended BSc Programme on a par with the Extended Degree Programmes in other faculties of the UFS.

The current entrance requirements[1] for the Extended BSc Programme are:

- a minimum AP score of 25 with a degree endorsement;
- Grade-12 NSC Mathematics at a minimum achievement Level 3;
- Grade-12 Physical Sciences at a minimum achievement Level 3 or Life Sciences at a minimum achievement Level 4; and
- the language of instruction at a Grade-12 minimum achievement Level 4.

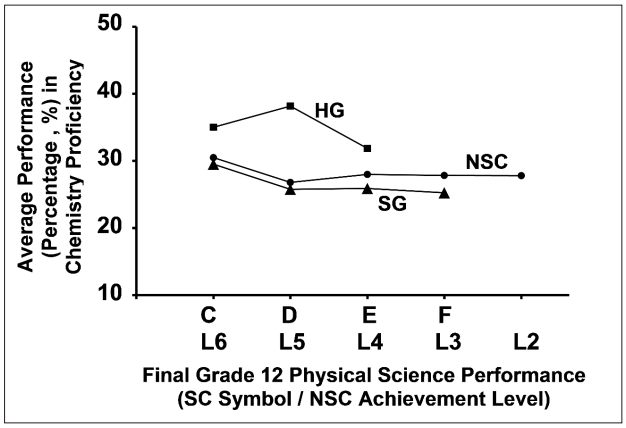
Currently, students in the Extended BSc Programme register in the first year of study for 8 developmental/foundation credits as well as 24 mainstream credits in Chemistry, 24 mainstream credits in Biology and 32 foundation Mathematics credits, similarly to that in the UPP.[2,4] Students also register for three development modules: Academic Language; Skills and Competencies for Life Long Learning: Natural Science; and Computer Literacy. An Extended BSc student who has passed all the modules of the first year of the programme can proceed to register for the second year of the programme on the Bloemfontein Campus.

Since the inception of the Extended BSc Programme in 2005 in the Faculty of Natural and Agricultural Science at the UFS, 2995 students have registered for the programme. From the available student data and the entrance requirements, student profiles regarding school achievement levels in Mathematics and Physical Sciences in relation to the AP score have been constructed. These student profiles indicate that some students achieved the required AP score for mainstream study but failed to achieve the required mathematics and science requirements. On the other hand, some students obtained the required mathematics and science achievement level for mainstream study but did not obtain the required AP score. However, the majority of the students, did not meet any of the entrance requirements for mainstream graduate studies.

Until 2012, students in the Faculty of Natural and Agricultural Sciences at the UFS wrote proficiency tests during the first week of their first year of study. These tests included a pre-calculus test[5], a general thinking skills test,

as well as Chemistry and Physics proficiency tests. The latter two were based on the expected knowledge students should have obtained during their Grade 11 and 12 Physical Sciences studies. Our interest, however, is in the students' performance in Chemistry and therefore we only considered the performance in the Chemistry proficiency test.

The Chemistry proficiency test, overall, yielded very poor results. Figure 1 shows the average performance of students in the Extended BSc Programme in the Chemistry proficiency test in relation to their final Grade-12 examination achievement in Physical Sciences. The average proficiency test mark is almost similar for the various school achievement levels. However, students with a Senior Certificate higher-grade school achievement in Physical Sciences performed marginally better in the Chemistry proficiency test, while the performances of students with a Senior Certificate standard-grade achievement and a NSC achievement level almost corresponded.



HG, higher grade Senior Certificate (SC); SG, standard grade SC; NSC, National Senior Certificate

Figure 1: Average performance of students in the Extended BSc Programme in the Chemistry proficiency test in relation to their achievement levels in Grade-12 Physical Sciences for the period 2005 to 2012.

The performance of students in the Chemistry modules within the Extended BSc Programme is of relevance in this study. In 2005 – the first year of the programme – 134 students registered for CEM108, a 32-credit Chemistry module. Of these students, 36.5% passed the CEM108 module. In 2006, the CEM108 module was split into two 16-credit modules, i.e. CHE104 (Inorganic and Analytical Chemistry) and CHE194 (Physical and Organic Chemistry). The performance of the 1149 students registered for these modules during 2006 to 2010 is presented in Figure 2. The pass rates for the CHE104 and CHE194 modules were 55.4% and 61.4%, respectively. The marginally better performance in CHE194 could be the result to the fact that CHE104 was a prerequisite for CHE194.

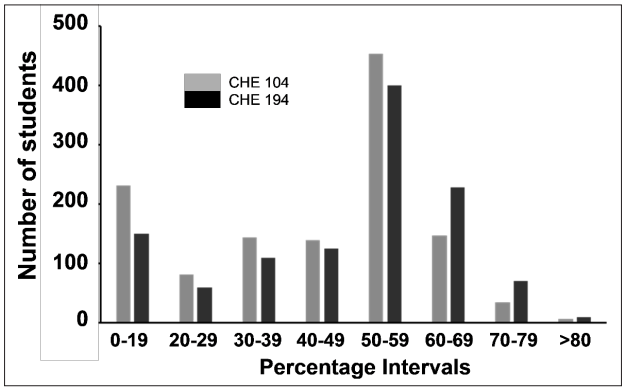


Figure 2: Performance of Extended BSc students in the first-year Chemistry modules CHE104 and CHE194 over the period 2006 to 2010.

In 2011, first-year Chemistry for the Extended BSc Programme was split into four modules: CHE112 (Introduction to Chemistry: A foundation module) and CHE132 (Organic Chemistry) in the first semester, and CHE122 (Physical Chemistry) and CHE142 (Inorganic and Analytical Chemistry) in the second semester. CHE112 contributed 8 foundation credits, whereas the other three each contributed 8 mainstream credits. Because of administrative restructuring at the UFS in 2014, the module codes were changed as follows: CHE112 to CHEM1552, CHE132 to CHEM1532, CHE122 to CHEM1622 and CHE142 to CHEM1642. The first semester CHEM1552 and CHEM1532 modules, as well as either the foundation Mathematics module, MATD1554, or an achievement Level 4 in Mathematics in the final Grade-12 NSC examination, are prerequisites for the second semester CHEM1622 module, whereas CHEM1532 is excluded as a prerequisite for CHEM1642. The students in the Extended BSc Programme performed satisfactorily in these first-year Chemistry modules (Figure 3). For the period 2011 to 2015, the pass rates for the first-semester CHEM1552 and CHEM1532 modules were 92.1% and 81.2%, respectively, and the pass rates for the second-semester CHEM1622 and CHEM1642 modules were 88.8% and 87.0%, respectively. It is interesting to note that the BSc Extended students performed better than the UPP students[4] in the first semester (92.1% vs 84% for CHEM1552; 81.2% vs 77.7% for CHEM1532), but that the UPP students were on par with the BSc Extended students in the second semester (87.9% vs 87.0% for CHEM1642; 91.1% vs 88.8% for CHEM1622).

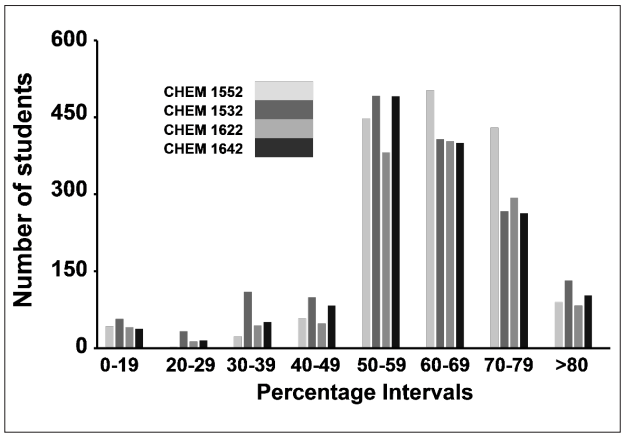


Figure 3: Performance of Extended BSc students in the first-year Chemistry modules CHEM1552, CHEM1532, CHEM1622 and CHEM1642 over the period 2011 to 2015.

Of the 2995 students in the Extended BSc programme during the period 2005 to 2015, 14.0% registered for at least one second-year Chemistry module. The reason why some students registered for only one Chemistry module can be ascribed to some degree programmes in the Faculty of Natural and Agricultural Sciences requiring only a specific Chemistry module. The performance of the Extended BSc students in the second-year Chemistry modules is given in Figure 4. The pass rates for the first-semester modules CHEM2614 (Physical Chemistry) and CHEM2632 (Analytical Chemistry) were 71.2% and 80.8%, respectively, and the pass rates for the second semester modules CHEM2642 (Inorganic Chemistry) and CHEM2624 (Organic Chemistry) were 89.0% and 57.6%, respectively.

Only 6.5% of the 2995 Extended BSc students over the sample period registered for at least one third-year Chemistry module. The pass rates in the first-semester third-year modules CHEM3714 (Analytical Chemistry) and CHEM3734 (Physical Chemistry) were 86.5% and 79.5%, respectively, and the pass rates for the second-semester third-year modules CHEM3724 (Inorganic Chemistry) and CHEM3744 (Organic Chemistry) were 68.4% and 72.4%, respectively. Figure 5 shows the performance of the students in the four third- year Chemistry modules.

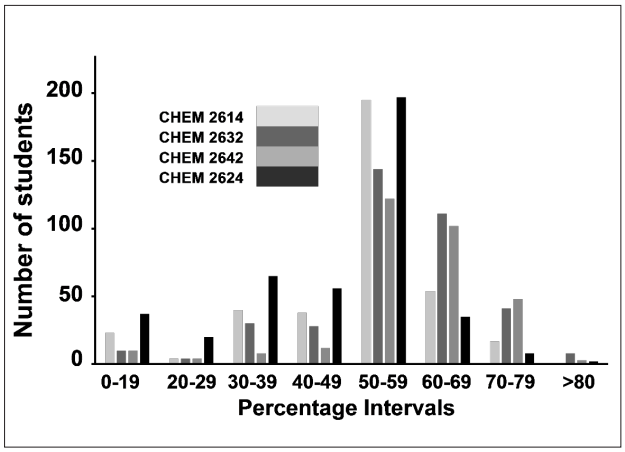


Figure 4: Performance of Extended BSc students in the second-year Chemistry modules CHEM2614, CHEM2632, CHEM2642 and CHEM2624 over the period 2006 to 2015.

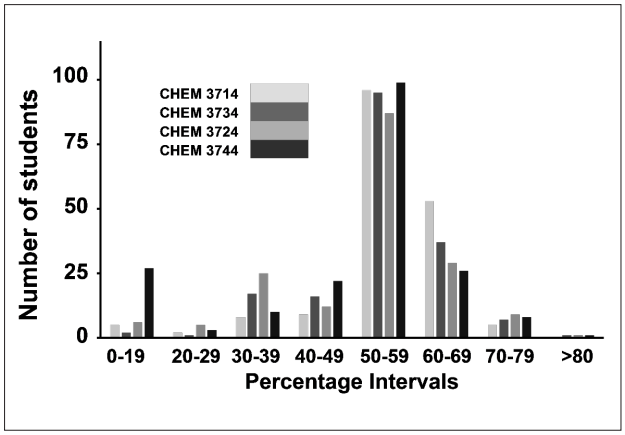


Figure 5: Performance of Extended BSc students in the third-year Chemistry modules CHEM3714, CHEM3734, CHEM3724 and CHEM3744 over the period 2007 to 2015.

The success of the programme is measured in terms of the number of students who graduate. Of the 2995 students, 20.3% (609 students) obtained a first degree over the sample period. Of these, 486 students (16.2%) obtained a first degree in the Faculty of Natural and Agricultural Sciences and 123 students (4.1%) registered in other faculties after the first year in the programme and obtained a first degree in those faculties. Table 1 shows the number of first degree graduates in the different study programmes. Of the graduates, 135 proceeded to an honours degree and 15 of those completed a master's degree. No student in this sample completed PhD study. A number of the graduates chose education as a career and proceeded to a Postgraduate Certificate in Education; 42 graduates (1.4%) obtained this postgraduate teaching certificate.

Our interest, however, is the performance in, and outcomes regarding Chemistry. The above-mentioned statistics show the performance of the students in the Extended BSc Programme in Chemistry over the past 10 years. Of the 2995 students in the programme during the sample period, 75 students (2.5%) obtained a BSc degree majoring in Chemistry (Table 1). Because only Chemistry had modules eligible for mainstream credits in the first year of the programme in the period 2005–2013, it was expected that the Chemistry course would benefit a great deal more from the programme. Factors like Mathematics prerequisites and facility constraints for Chemistry had a direct effect on the number of Chemistry graduates.

Of the 75 Chemistry graduates, only 11 students proceeded with Chemistry and obtained an honours degree in Chemistry. Four of these honours graduates proceeded towards a MSC degree in Chemistry.

Table 1 : Number of first degrees conferred on students who started in the Extended BSc Programme of the University of the Free State since its inception

Degree programme	Number of degrees
Natural and Agricultural Sciences	
Agriculture	44
Botany	38
Building Science	4
Chemistry	75
Consumer Science	3
Physics	4
Genetics	38
Geology	55
Geography	50
Information Technology	36
Mathematics/Statistics	12
Medical Microbiology/Human Biology	20
Microbiology/Biochemistry	73
Zoology	34
Other faculties	
Commerce	16
Education	8
Humanities	78
Medicine	4
Social Sciences	17

The Chemistry department and the Faculty of Natural and Agricultural Sciences of the UFS are proud of what has been achieved with the Extended BSc Programme. The fact that 609 first degrees and 165 postgraduate degrees and certificates have been conferred on students who started in the Extended BSc Programme since inception of the programme, underlines the success of the Extended BSc Programme of the UFS.

The Extended Degree Programmes of the UFS offer an opportunity to students who, as a consequence of personal challenges and also limitations in the school system, did not meet the entrance requirements to graduate studies, to proceed into graduate study after a one-year access programme. Congratulations are in order to the staff of the UFS who developed the Extended Degree Access Programmes, for providing the youth of South Africa with an opportunity to build themselves a promising career.

Acknowledgement

We thank Annemari Eksteen and Jacques Botha of the Directorate for Institutional Research and Academic Planning at the UFS for their help with the compilation of the data.

References

1. University of the Free State. Faculty of Natural and Agricultural Sciences: Yearbooks [homepage on the Internet]. No date [cited 2017 May 28]. Available from: http://www.ufs.ac.za/natagri/faculty-of-natural-and-agricultural-sciences-home/academic-information/yearbooks

2. Hay HR, Marais F. Bridging programmes: Gain, pain or all in vain. S Afr J High Educ. 2004;18(2):59–75. http://dx.doi.org/10.4314/sajhe.v18i2.25454

3. Rabie NE. Access to higher education: The case of the Career Preparation Programme at the University of the Free State [mini thesis]. Cape Town: University of the Western Cape; 2008.

4. Marais C, Meintjes R, Dennis CR, Swarts JC. Performance in Chemistry of students who started in the University Preparation Programme: The ripple effect. S Afr J Sci. 2016;112(7/8), Art. #a0162, 3 pages. http://dx.doi.org/10.17159/sajs.2016/a0162

5. Dennis CR, Murray DM. Success in first-year mathematics: School-leaving examinations and first-year performance. S Afr J Sci. 2012;108(7/8), Art. #1325, 3 pages. http://dx.doi.org/10.4102/sajs.v108i7/8.1325

Towards an integrated ecological restoration approach for abandoned agricultural fields in renosterveld, South Africa

AUTHOR:
Sheunesu Ruwanza[1]

AFFILIATION:
[1]Department of Ecology and Resource Management, School of Environmental Science, University of Venda, Thohoyandou, South Africa

CORRESPONDENCE TO:
Sheunesu Ruwanza

EMAIL:
ruwanza@yahoo.com

KEYWORDS:
biodiversity; conservation; Fynbos Biome; land abandonment; ecological recovery; succession

HOW TO CITE:
Ruwanza S. Towards an integrated ecological restoration approach for abandoned agricultural fields in renosterveld, South Africa. S Afr J Sci. 2017;113(9/10), Art. #a0228, 4 pages. http://dx.doi.org/10.17159/sajs.2017/a0228

The importance of intact renosterveld vegetation to the maintenance of ecosystem health and services has been increasingly highlighted in the literature.[1-3] Similarly, acknowledgement of heavy transformation (over 80%) and fragmentation (approximately 5% intact remnant patches remain) of renosterveld vegetation has been well documented.[4,5] Renosterveld – a shrubby vegetation type dominated by *Elytropappys rhinocerotis* (L.f.) Less – has been heavily transformed and replaced by agriculture, mostly vine, olive and wheat cultivation.[6] The fragmented renosterveld remnants remain under threat of being cleared for new agricultural lands or being subjected to disturbances like overgrazing, fire and invasion by alien plants.[5] There is, therefore, the need for conservation of renosterveld through securing the existing remnants, linking fragments through ecological restoration and spreading awareness about good management practices.

Ecological restoration of abandoned agricultural fields using either the successional or alternative state models creates an opportunity to increase renosterveld size and restore ecosystem function and structure. Ecological restoration is dependent on the type, size, disturbance history and survival of introduced species (where active restoration is initiated). The key questions linked to ecological restoration pertain to when and how to implement ecological restoration intervention. The answers to these questions vary given the many angles and components of the questions, e.g. have biotic and abiotic ecological thresholds been passed or have ecological, social and economic concerns been considered?[7,8]

There have been a few articles over the past 10 years that have addressed ecological restoration in renosterveld. These articles are a product of a few field-based research experiments that have been conducted on a small scale and their aim is to find an appropriate restoration technique that is cost effective. This commentary presents an evaluation of field-based research experiments on ecological restoration in renosterveld abandoned agricultural fields. Using 11 articles published between 2005 and 2016, three important ecological restoration themes were identified that justify some discussion. These themes are related to (1) factors hindering ecological restoration in renosterveld abandoned agricultural fields, (2) evaluating ecological restoration success in renosterveld abandoned agricultural fields and (3) moving towards an integrated ecological restoration approach for abandoned agricultural fields in renosterveld. The last theme is an outcome of the evaluation process aimed at developing a new ecological restoration approach that can be used to achieve restoration success in abandoned agricultural fields of renosterveld.

Factors hindering ecological restoration

Previous field-based research experiments have identified several factors that impede ecological restoration in renosterveld abandoned agricultural fields. The factors can be grouped into three broader groups: vegetation factors, soil and environmental conditions (Table 1). Vegetative factors, such as seed sources, competition from invasive alien grasses and predation of recruiting native plants by animals have been identified as some of the factors that hinder ecological restoration in renosterveld. Previous studies on soil-stored seed bank in renosterveld abandoned agricultural fields have reported a lack of native species in soil-stored seed banks.[9,10] The lack of soil-stored seed banks in renosterveld abandoned agricultural fields is a result of previous cultivation that has depleted the soil seed bank. An examination of seed dispersal has shown that dispersal of native species into abandoned agricultural fields does occur, but seeds struggle to germinate on arrival.[11,12] In areas where grazing is allowed, the few native species that manage to recruit are grazed by animals whilst they are still young.[6] If not grazed, environmental factors, e.g. the hot summer temperatures accompanied with the lack of rainfall, seem to affect seedling establishment.

Table 1: Summary of factors that hinder ecological restoration success, based on the 11 reviewed field-based ecological restoration experiments conducted in abandoned agricultural fields of renosterveld

Vegetation	Soil	Environmental
Seed source	Elevated soil nutrients	Climatic patterns
Predatory	Competition	
Invasive alien grasses		

The proliferation of invasive alien grasses, which is linked to high soil nutrients in abandoned agricultural fields of renosterveld, has also been identified to hinder ecological restoration.[13,14] Past fertilisation has been identified as one of the drivers of the observed high soil nutrients. Studies on recovery following alien grass removal have concluded that most alien grass control options (e.g. herbicide application) produce successful results, but they negatively affect native species recovery.[13,14] Also, a study on soil nutrient manipulation has reported that the technique is an ineffective ecological restoration method.[14] Both invasive alien grasses and the high soil nutrients create positive and negative feedback mechanisms that hinder plant and soil recovery.[15,16] For example, nutrient-rich soils found in abandoned agricultural fields tend to facilitate the growth of fast-growing invasive alien grasses, which when dead contribute to high soil nutrient content. This positive and negative feedback mechanism tends

to contribute to the maintenance of a degraded abandoned agricultural field.[16] Competition for resources (e.g. water and soil nutrients) has been highlighted as a factor hindering ecological restoration in renosterveld.[6,10] Competition is linked to the proliferation of fast-growing invasive grasses that are known to utilise nutrient resources faster than native species.[14] Most environmental factors that hinder ecological restoration have not yet been studied in renosterveld, but have been observed during some experiments. For example, the mortality rate for most introduced native species was high in summer because of a lack of water and high temperatures associated with summer in the Western Cape Province.[14]

Evaluation of ecological restoration success

Assessing the success of any ecological restoration project is important so as to justify restoration as a management intervention that improves the provision of ecosystem services.[17] Evaluating ecological restoration is not straightforward given the debate around how best to measure success.[17] If ecological restoration is defined as 'the process of assisting the recovery of an ecosystem that has been degraded, damaged, or destroyed', then restoration success should be guided by three outcomes: recovery of vegetation structure, species diversity and abundance and ecological processes.[17,18] This implies that recovery of abandoned agricultural fields in renosterveld should entail the recovery of diverse native species and ecological processes, e.g. physical (soil structure), chemical (soil nutrients) and biological (soil bacteria) processes.

Based on previous field-based research experiments in renosterveld, three conclusions can be made with regard to restoration success. Firstly, studies in which native species were introduced in abandoned agricultural fields of renosterveld have concluded that the introduction of native species produces little ecological restoration success (Table 2).[6,14] However, a recent study has shown that seeding has the potential to yield positive ecological restoration results if combined with other grass

Table 2: Summary of research aims, results and outcomes of ecological restoration successes, based on the 11 reviewed field-based ecological restoration experiments conducted in abandoned agricultural fields of renosterveld

Broader restoration aims	Restoration results	Restoration outcomes	Reference
Influence of competition and herbivory on old field restoration	Introduced native plants competed with alien grasses for resources. Competition negatively affected germination and seedling survival and growth.	Competition from grasses and herbivory need to be reduced for restoration to be successful.	Midoko-Iponga et al.[6]
Effective alien grass control linked to old field restoration success	Controlling alien grasses by burning is cheaper but affected native species seed banks. It also promoted alien grass infestations. The light burn of invasive grasses stimulated plant recovery. Mowing was cheaper but triggered little species recovery. Hand pulling and herbicide application are expensive methods. Herbicide application inhibited native species recovery.	Integrated control methods are necessary to effectively control and restore native species.	Musil et al.[13]
Restoration implication of seed dispersal in the dung of herbivores	Alien grasses dominated seedlings from herbivore dung. Only one shrub species was identified in herbivore dung.	The presence of large herbivores in old fields could retard recovery because they disperse seeds of alien grasses.	Shiponeni and Milton[12]
Hydrological and soil retention services are benefits of restored renosterveld	Infiltration was higher in renosterveld than in transformed renosterveld. Intact renosterveld reduced wind speed and aeolian loads compared to transformed renosterveld.	Restoring renosterveld old fields has some ecological benefits.	O'Farrell et al.[20]
Vegetation recovery in abandoned renosterveld croplands	Therophyte and chamaephyte species increased with an increase in abandonment years. Species richness increased with age since cropland abandonment.	Vegetation recovery in abandoned croplands is occurring naturally but the recovery rate differs amongst the life forms.	Van der Merwe and van Rooyen[19]
Introducing pioneer species and bush clumps for restoration purposes	Introduced plants had high germination rates in the greenhouse but low germination and establish rates under field conditions. Bush clumps increased seed dispersal but germination and established rates following seed dispersal was low.	Restoration using early succession species and natural dispersal vectors does not trigger native vegetation recovery.	Heelemann et al.[11]
Effects of soil manipulation using sucrose on plant and soil recovery	Sucrose addition negatively affected the growth of both native and alien grasses. Sucrose addition had little effect on bacterial activity.	Soil manipulation by adding sucrose is ineffective as a restoration option because it negatively affects the growth of both native and alien species.	Ruwanza et al.[14]
Seed bank of remnant and degraded renosterveld	Old fields were dominated by alien grasses, nutrient-rich soils and depleted indigenous soil seed banks. Seed smoking did not influence seedling recruitment.	Abandoned agricultural fields have low restoration potential because of the lack of an indigenous soil seed bank.	Heelemann et al.[9]
Carbon sequestration and restoration through fallowing	Ecosystem carbon stocks in fallow fields were equal to those in intact renosterveld. Fallowing had the potential to sequestrate carbon.	Carbon financing could be used to promote restoration of old fields.	Mills et al.[21]
Impacts of herbivores and fire on renosterveld vegetation	Herbivory, fire and their interaction negatively affected vegetation composition in old fields. Burning and grazing caused the growth of unpalatable species, whilst grazing restrictions caused the growth of palatable species.	Herbivory removal can cause recovery of palatable species. Burning combined with restricted grazing in old fields causes proliferation of alien grasses.	Radloff et al.[4]
Seeding recruitment under multifactorial restoration treatments	Seeding alone was ineffective, and needed to be combined with another method, e.g. herbicide application to remove grasses, burning to enhance seed germination, tillage for soil preparation and grass removal, or rodent exclusion.	Seeding does improve native species presence in old fields but needs to be combined with other restoration methods.	Waller et al.[10]

removal techniques.[10] Secondly, a study on natural recovery concluded that vegetation recovery is occurring naturally in mountain renosterveld (Table 2).[4,19] Lastly, studies on restoration benefits have confirmed that restoring abandoned agricultural fields improves hydrological and soil retention services.[20,21] The above three outcomes are an indication that ecological restoration initiatives in renosterveld are yielding mixed results. The slow recovery reported in some studies could be a result of the previously identified factors that hinder ecological restoration. Indeed, abandoned agricultural fields whose restoration is constrained by the identified factors represent old fields that are in a persistent degraded state. In such state, both biotic and abiotic factors have been strongly altered by the previous cultivation.[16]

The applicability of natural succession as an ecological restoration initiative in lowland renosterveld might be difficult and may take several years, although it is yielding positive ecological restoration results in mountain renosterveld.[19] Firstly, there exists a possibility that natural succession in lowland renosterveld may trigger bush encroachment. This is possible where grazing and browsing animals are allowed to utilise palatable species, leaving unpalatable species to dominate abandoned agricultural fields.[4] Besides, grazing has also been reported to disperse seeds of invasive alien grasses rather than those of native shrubs.[12] Secondly, natural succession will only be successful where remnants of renosterveld are close to the abandoned agricultural fields so as to allow seed dispersal to take place. Seed dispersal is reported to be taking place, but is unfortunately spreading invasive alien grasses.[11,12]

Towards an integrated ecological restoration approach

Restoration attempts in abandoned agricultural fields of renosterveld should move towards an integrated ecological restoration approach (Figure 1). The integrated ecological restoration approach should seek to combine several restoration techniques aimed at restoring abandoned agricultural fields. The advantage of adopting an integrated ecological restoration approach is that it will allow several methods to be implemented consecutively or combined. Integrating several ecological restoration methods has yielded positive results in old fields in Virginia (USA), where herbicide application combined with soil nutrient manipulation and seeding facilitated native grass recovery.[22] However, the integrated restoration approach has its own challenges. Firstly, the selection of an appropriate ecological restoration technique may be difficult and require several trials. Secondly, different restoration approaches may yield different results that may positively or negatively affect other approaches. For example, alien grass mowing might result in cut grass biomass blocking sunlight penetration and reducing seed-to-soil contact resulting in reduced native species germination. Lastly, integrating different ecological restoration approaches requires constant monitoring and evaluation, which is likely to be time consuming and costly.

A successfully integrated restoration approach for renosterveld must include four phases: control of invasive grasses, recovery of soil properties, recovery of vegetation and monitoring and evaluation (Figure 1). The control of invasive grasses can be achieved by using different techniques aimed at grass removal, e.g. fire, herbicide application, hand pulling or mowing.[13] Measures to restore soil properties – e.g. soil nutrient manipulation, top soil removal and tillage – should aim to recover soil physico-chemical properties. Measures to recover native species (e.g. seed sowing, seedling transplanting and perching) should aim to restore vegetation diversity, structure and function. The monitoring and evaluation process should concentrate on both soil and plant recovery (Figure 1). In order for the suggested integrated ecological restoration approach to be cost-effective, soil transfer which aims for both soil and plant recovery simultaneously can be tried. After implementing grass control measures, soil transfer from renosterveld remnant areas to abandoned agricultural fields has the advantages of improving soil physico-chemical and microbial properties and processes.[22,23] It also introduces intended natural species that are present in soil flora transferred from renosterveld remnant areas. However, soil transfer has its own negative impacts, e.g. the transfer of unwanted species, movement of large quantities of soil[23], it is costly and the soil is disturbed.[24] Besides these challenges, several studies have reported that soil transfer provides good restoration results.[23-25]

In conclusion, previous studies have shown that ecological restoration in abandoned agricultural fields of renosterveld is producing mixed results.

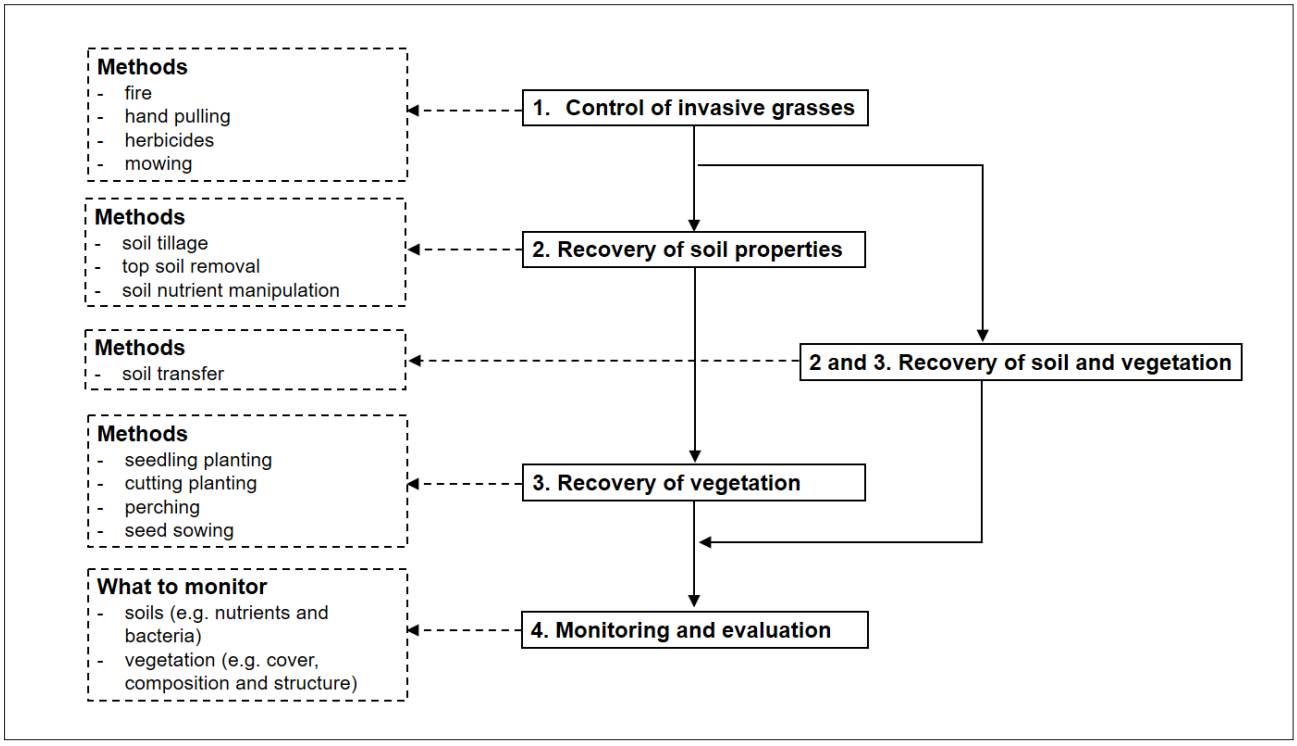


Figure 1: The conceptual framework of the suggested integrated ecological restoration approach for renosterveld.

One can generally conclude that it is on a slow recovery trajectory. Factors inhibiting ecological restoration are multiple and seem to interact in a way that maintains the degraded state. An integrated ecological restoration approach in which multiple ecological restoration methods are implemented should be adopted for ecological restoration success to be realised in renosterveld. For the integrated ecological restoration approach to be economical, soil transfer from intact renosterveld areas to abandoned agricultural fields can be adopted following implementation of invasive alien grass control techniques.

References

1. Cowling RM, Pierce SM, Moll EJ. Conservation and utilisation of South Coast renosterveld, an endangered South African vegetation type. Biol Conserv. 1986;37:363–377. http://dx.doi.org/10.1016/0006-3207(86)90078-9
2. Rebelo T. Renosterveld: Conservation and research. In: Low AB, Jones FE, editors. The sustainable use and management of renosterveld remnants in the Cape Floristic Region. Proceedings of a Symposium. FCC Report 1995/4. Cape Town: Flora Conservation Committee, Botanical Society of South Africa; 1995. p. 32–42.
3. Bergh NG, Verboom GA, Rouget M, Cowling RM. Vegetation types of the Greater Cape Floristic Region. In: Allsopp N, Colville JF, Verboom GA, editors. Fynbos: Ecology, evolution, and conservation of a megadiverse region. Oxford: Oxford University Press; 2014. p. 26–46. http://dx.doi.org/10.1093/acprof:oso/9780199679584.003.0001
4. Radloff FGT, Mucina L, Snyman D. The impact of native large herbivores and fire on the vegetation dynamics in the Cape renosterveld shrublands of South Africa: Insights from a six-yr field experiment. Appl Veg Sci. 2014;17:456–469. http://dx.doi.org/10.1111/avsc.12086
5. Kemper J, Cowling RM, Richardson DM. Fragmentation of South African renosterveld shrublands: Effects on plant community structure and conservation implications. Biol Conserv. 1999;90:103–111. http://dx.doi.org/org/10.1016/S0006-3207(99)00021-X
6. Midoko-Iponga D, Krug CB, Milton SJ. Competition and herbivory influence growth and survival of shrubs on old fields: Implications for restoration of renosterveld shrubland. J Veg Sci. 2005;16:658–692. http://dx.doi.org/10.1111/j.1654-1103.2005.tb02411.x
7. Esler KJ, Holmes PM, Richardson DM, Witkowski ETF. Riparian vegetation management in landscapes invaded by alien plants: Insights from South Africa. S Afr J Bot. 2008;74:401–552. http://dx.doi.org/org/10.1016/j.sajb.2008.01.168
8. Aronson J, Milton SJ, Blignaut JN. Restoring natural capital: Science, business, and practice. Washington: Island Press; 2007.
9. Heelemann S, Krug CB, Esler KJ, Reisch C, Poschlod P. Soil seed banks of remnant and degraded Swartland Shale Renosterveld. Appl Veg Sci. 2013;16:585–597. http://dx.doi.org/10.1111/avsc.12026
10. Waller PA, Anderson PML, Holmes PM, Allsopp N. Seedling recruitment responses to interventions in seed-based ecological restoration of Peninsula Shale Renosterveld, Cape Town. S Afr J Bot. 2016;103:193–209. http://dx.doi.org/org/10.1016/j.sajb.2015.09.009
11. Heelemann S, Krug CB, Esler KJ, Reisch C, Poschlod P. Pioneers and perches-promising restoration methods for degraded renosterveld habitats? Restor Ecol. 2012;20(1):18–23. http://dx.doi.org/10.1111/j.1526-100X.2011.00842.x
12. Shiponeni NN, Milton SJ. Seed dispersal in the dung of large herbivores: Implications for restoration of Renosterveld shrubland old fields. Biodivers Conserv. 2006;15:3161–3175. http://dx.doi.org/10.1007/s10531-005-6317-5
13. Musil CF, Milton SJ, Davis GW. The threat of alien invasive grasses to lowland Cape floral diversity: An empirical appraisal of the effectiveness of practical control strategies. S Afr J Sci. 2005;101:337–344.
14. Ruwanza S, Musil CF, Esler KJ. Sucrose application is ineffectual as a restoration aid in a transformed southern African lowland fynbos ecosystem. S Afr J Bot. 2012;80:1–8. http://dx.doi.org/org/10.1016/j.sajb.2012.01.009
15. Suding KN, Gross KL, Houseman GR. Alternative states and positive feedbacks in restoration ecology. Trends Ecol Evol. 2004;19:46–53. http://dx.doi.org/org/10.1016/j.tree.2003.10.005
16. Cramer VA, Hobbs RJ, Standish RJ. What's new about old fields? Land abandonment and ecosystem assembly. Trends Ecol Evol. 2008;23(2):104–112. http://dx.doi.org/org/10.1016/j.tree.2007.10.005
17. Wortley L, Hero J-M, Howes M. Evaluating ecological restoration success: A review of the literature. Restor Ecol. 2013;21(5):537–543. http://dx.doi.org/10.1111/rec.12028
18. Society for Ecological Restoration (SER). Society for Ecological Restoration international's primer of ecological restoration [homepage on the Internet]. c2004 [cited 2017 Apr 04]. Available from: http://www.ser.org/resources/resources-detail-view/ser-internationalprimer-on-ecological-restoration
19. Van der Merwe H, Van Rooyen MW. Life form and species diversity on abandoned croplands, Roggeveld, South Africa. Afr J Range For Sci. 2011;28(2):99–110. http://dx.doi.org/10.2989/10220119.2011.642097
20. O'Farrell PJ, Donaldson JS, Hoffman MT. Local benefits of retaining natural vegetation for soil retention and hydrological services. S Afr J Bot. 2009;75:573–583. http://dx.doi.org/org/10.1016/j.sajb.2009.06.008
21. Mills AJ, Birch S-JC, Stanway R, Huyser O, Chisholm RA, Sirami C, et al. Sequestering carbon and restoring renosterveld through fallowing: A practical conservation approach for the Overberg, Cape Floristic Region, South Africa. Conserv Lett. 2013;6:255–263. http://dx.doi.org/10.1111/conl.12003
22. Priest A, Epstein H. Native grass restoration in Virginia old fields. Castanea. 2011;76(2):149–156. https://doi.org/10.2179/09-056.1
23. Hölzel N, Otte A. Restoration of a species-rich flood meadow by topsoil removal and diaspore transfer with plant material. Appl Veg Sci. 2003;6:131–140. http://dx.doi.org/10.1111/j.1654-109X.2003.tb00573.x
24. Bulot A, Potard K, Bureau F, Bérard A, Dutoit T. Ecological restoration by soil transfer: Impacts on restored soil profiles and topsoil functions. Restor Ecol. 2017;25:354–366. http://dx.doi.org/10.1111/rec.12424.
25. Muller I, Buisson E, Mouronval J-B, Mesléard F. Temporary wetland restoration after rice cultivation: Is soil transfer required for aquatic plant colonization? Knowl Manag Aquat Ecosyst. 2013;411(3):1–17. https://doi.org/10.1051/kmae/2013067

Pollination: Impact, role-players, interactions and study – A South African perspective

AUTHORS:
Annemarie Gous[1,2]
Sandi Willows-Munro[2]
Connal Eardley[2,3]
Zacharias H. (Dirk) Swanevelder[1,4]

AFFILIATIONS:
[1]Biotechnology Platform, Agricultural Research Council, Onderstepoort, Pretoria, South Africa
[2]School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa
[3]Plant Protection Research Institute, Agricultural Research Council, Pretoria, South Africa
[4]College of Agriculture and Environmental Sciences, University of South Africa, Pretoria, South Africa

CORRESPONDENCE TO:
Dirk Swanevelder

EMAIL:
swanevelderd@arc.agric.za

DATES:
Received: 24 Oct. 2016
Revised: 23 Jan. 2017
Accepted: 16 May 2017

KEYWORDS:
floral choice; bees; plant–pollinator interactions; metabarcoding; pollen

HOW TO CITE:
Gous A, Willows-Munro S, Eardley C, Swanevelder ZH. Pollination: Impact, role-players, interactions and study – A South African perspective. S Afr J Sci. 2017;113(9/10), Art. #2016-0303, 8 pages. http://dx.doi.org/10.17159/sajs.2017/20160303

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Belgian Directorate-General for Development Cooperation, Partim GTI; National Research Foundation (South Africa)

Plant–pollinator interactions are essential for maintaining both pollinator and plant communities in native and agricultural environments. Animal-instigated pollination can be complex. Plants are usually visited by a number of different animal species, which in turn may visit flowers of several plant species. Therefore, the identification of the pollen carried by flower visitors is an essential first step in pollination biology. The skill and time required to identify pollen based on structure and morphology has been a major stumbling block in this field. Advances in the genetic analysis of DNA, using DNA barcoding, extracted directly from pollen offers an innovative alternative to traditional methods of pollen identification. This technique, which is reviewed in detail, can be used on pollen loads sampled from bees in the field and from specimens in historic collections. Here the importance of pollination, the role-players involved, their management and the evolution of their interactions, behaviour and morphology are reviewed – with a special focus on South African bees.

Significance:

- Pollen metabarcoding will enable the identification of pollen for a multitude of uses, including agriculture, conservation and forensics.
- Plant–pollinator interaction documentation through pollen identification gives a more certain record of a visitor being a pollinator rather than a flower visitor that could be a nectar gatherer.

Introduction

South Africa has one of the world's most diverse landscapes, with high plant and pollinator diversity and endemism.[1,2] Healthy plant–pollinator interactions are important to maintain both native plant and pollinator communities. The interactions between pollinators and their host plants are complex and very little is currently known about the floral choices of indigenous bees, specifically in South Africa. As a result of pressure from urban development, overexploitation of natural resources, and climate change, many of the nine biomes in South Africa are under threat.[3] The effects of an anthropogenic influence on the environment dictates improved methods of studying plant–pollinator interactions, to understand how they may be influenced by environmental and ecological changes.

Most plant–pollinator interaction studies rely on lengthy field-based experiments.[2] Observing pollinators in the field is not the only way to study their interactions with plants. Pollinator pollen load identification allows insight into the species' floral visits. Pollen loads provide a snapshot of the interactions with the plant community at the time they were caught. Bees (Hymenoptera: Apoidea) caught at a flower patch, for example, can have varying pollen loads because they can be either on their way to the flowers from the nest, or on their way from the flowers to the nest, or busy foraging at the patch. Neither identifying the pollen found on a sampled bee, nor netting a bee on a flower, can give definite answers regarding the plants it pollinates, and similarly, inferences on possible fruit and seed set cannot be made.[4] Some pollination inferences can nonetheless be drawn, especially if multiple bee samples are investigated. Pollen loads sampled from pollinators could be identified by classic microscopic palynology or genetic methods. These two approaches are discussed in detail below.

Here, the role of pollination in agriculture and natural plant populations, with a special focus on bee pollination, is elucidated. The value of bee pollination in a South African context is reviewed and discussed, as is the potential impact that the introduction of a foreign bee species could have on highly diverse native bee populations. Advances in studying plant–pollinator interactions using genetic methods are also reviewed.

The value of pollination

Functional ecosystems require various essential ecosystem services to be performed. Ecosystem services are defined as services provided to humans by organisms that interact in the ecosystem and pollination is one such extremely important service.[5] Plant–pollinator interaction is, in most instances, an intimate mutualistic relationship, in which both parties are reliant on each other for survival – plants for reproduction and pollinators for food or other forms of reward. Although a plant might have multiple pollinators, it is possible that one or more of these pollinators are specialists and may, therefore, rely heavily on that specific plant taxon for survival.[6] A decline in the host plant numbers would ultimately lead to a decline in its specialist pollinators, and vice versa, with important impacts on maintaining biodiversity and ecosystems.

Pollination is extremely important, not only in natural ecosystems but also in artificial production environments. Biotic pollination of crops is important to consider from an agricultural production perspective because approximately one-third of all human food consumption results from animal-pollinated plants, of which up to 75% is used directly as food.[6] A decline in crop pollinator populations would thus negatively impact crop production. The importance of pollinators has been illustrated in a study conducted on 137 single crops and five commodities, in which increases of 68.4% in production of the leading single crops and 71.6% in production of commodity crops were found with animal pollination.[6] It was estimated that native insects in the United States of America (excluding the introduced honeybee) were solely responsible for USD3.07 billion worth of fruit and vegetable production in 2003.[5] This figure clearly indicates the worth of indigenous pollinators to society for maximal crop production.

Plant–pollinator interactions

Pollination is an important ecosystem service and often crucial for the survival of both parties involved, but is, in essence, an inadvertent process. From the view of the pollinator, it is not its specific goal to provide this service to its mutualistic plant partner, but it is rather a coincidental result of its actions while visiting these hosts.[7] Bees, for example, can deliberately collect pollen, or pollen can passively adhere to bee bodies while they are visiting flowers.[8] Mutualism between the plant and the pollinator is based on rewards (pollen, nectar and oil) that the pollinator receives from the plant, and the plant gains the service of successful pollination, thereby securing its reproductive success.[7-9] This mutualism can be facultative or obligate, depending on whether the plant is self-compatible or whether it is monoecious or dioecious.[7] Parasitic interactions are also possible when a potential pollinator takes pollen and nectar from the plant without playing any role in its pollination.[7,8] Interactions between plants and potential pollinators therefore range from parasitic to obligate mutualistic, with each plant–pollinator interaction a developing relationship, based on how the plant and pollinator adapt to suit each other.

Evolution of plant–pollinator interactions

Interactions between pollinators and their target plants are usually regarded as being either generalised or specialised.[10] Generalisation describes an interaction in which a plant has flowers that are morphologically accessible and attractive to many different pollinator species. Specialisation refers to flowers that are sufficiently specialised as to be attractive and/or accessible to only a single type of pollinator.[10,11] The same can be applied from the pollinator's perspective – that is, the range of plants a pollinator prefers relates to it being generalised or specialised, as discussed below.[10,12] In some extreme cases, both parties could co-evolve morphologically and behaviourally to allow only one-on-one plant–pollinator interactions, whereby the plant protects access to rewards for its specific pollinator – a feature of, amongst others, many genera in the Orchidaceae.[13]

It has been argued that the formation of specific floral structures in plants is largely driven by means of natural selection from their respective pollinators.[9] Better pollinator-flower compatibility would, therefore, result in higher selection, through increased fertilisation events of these individuals. The pollinator-plant interaction is important as pollinator or floral adaptations can drive speciation as suggested by the diverse floras of the Cape region, South Africa.[14] Flowers of angiosperms can gain suites of adaptive traits to make them more suitable to a certain type of pollinator or pollinator guild, also known as floral syndromes.[15] These floral adaptations or syndromes can lead to reproductive isolation and drive speciation, but isolation and speciation are not necessarily coupled.[16] It is important to remember that floral syndromes are not an absolute definition of a plant's pollinators but rather a description of how unrelated plants have evolved similar floral traits.[16]

Bees as pollinators

The most common and invariably important biotic pollinators of angiosperms are bees[6], as they actively collect pollen as food for themselves and/or their larvae[8]. There is an estimated 25 000 bee species in the world of which approximately two-thirds are taxonomically described.[9] Bees are important pollinators of tropical forest trees[17] and play an essential role in the pollination of smaller trees, shrubs and herbaceous plants[9]. Many crop plants are bee-pollinated, such as sunflower, tomato, canola, cowpea and coffee.[6,18]

South Africa has high levels of both bee and plant diversity, especially in the southern and western parts of the country that experience predominantly winter rainfall. Approximately 50% of the bee species known to occur in sub-Saharan Africa are also located in South Africa.[19] Moreover, 95% of the bee species found in the winter rainfall region occur only in southern Africa.[20] The moist, eastern part of the country has also been shown to be diverse in its bee species composition, albeit less so than in the arid western part of South Africa. However, the eastern region still contains a high endemism level of 75%.[19,20]

Although South Africa is particularly rich in pollinators and floristic diversity, relatively few comprehensive studies have investigated pollinator-plant interactions in the country as a whole. In one such study that focused on bees in the arid western region of southern Africa, 16 229 plants were visited by 924 species of non-*Apis* bees, wasps and pollen wasps.[2] The bees in this study were represented by 420 different species that visited 34 out of the 36 available plant families in the study area. The four plant families most frequently visited by bees were the Fabaceae, Asteraceae, Aizoaceae and Zygophyllaceae. The foraging habits of all the different bee families studied (i.e. the Colletidae, Andrenidae, Halictidae, Mellitidae, Megachilidae and Apidae) ranged from oligolectic (specialised) to narrowly or broadly polylectic (generalised) – although none of the families were nearly as polylectic as *Apis mellifera*. However, in the Mellitidae, half of the observed species were found to be oligolectic for members of the *Wahlenbergia* plant genus (Campanulaceae). This work has shed light on the diversity and foraging habits of pollinators in the arid and semi-arid regions of the western part of southern Africa.[2]

No similar regional-scale study of bees in the eastern part of South Africa is currently available. However, there are a few smaller studies of pollinators for specific plant species, for example, in the KwaZulu-Natal midlands in the eastern part of South Africa, *Wahlenbergia* were visited by halictid species and *A. mellifera*.[21] It is noteworthy that halictid bees have also been reported to be oligolectic for *Wahlenbergia* in Australia.[22] This report indicates a possible adaptation of bees to their locally available flora as was previously reported.[23] The few available studies, together with the high floristic and bee diversity and endemism in South Africa, highlight the need for further studies into the diversity of bee interactions with plant species – studies that are needed to elucidate floral choice patterns within South African bee populations.

Bee adaptations for foraging

Bees are active foragers, collecting various substances from flowers. During foraging activity, pollen grains become attached to their bodies. Specialised branched hairs on their bodies trap the pollen grains during the collection of pollen, nectar or oil.[24] Electrostatic charges on the hairs also aid in the transfer of pollen from anthers to the bee body. In addition, modifications of hairs on the mouthparts, undersides of the heads, or faces of bees all assist in extraction of pollen from flowers.[8,24] During foraging the pollen is groomed from the insect body into structures used to carry it to the nest. These transport structures, known as scopae, are brushes of hairs located on the hind tarsi of most bee species, or on the bottom of the abdomen as in the Megachilidae.[8] Structural and behavioural adaptations for the collection of pollen have previously been reviewed in more detail.[24] When pollen is groomed into transport structures, these pollen grains are generally not available for pollination[25] as they are tightly packed. Loosely adhered pollen grains on bee bodies are more important in pollination.

Generalist versus specialist interactions

Bees can exhibit generalist or specialist behaviour in their floral choices for specific requirements. Most bees within the eusocial groups – such as honeybees, bumblebees and most stingless bees – are polylectic in terms of pollen and nectar collection.[8] They visit plants from a wide variety of taxa that are available for pollen and nectar collection. Polylectic bees still show floral constancy, that is, they make repetitive visits to plants from the same taxon that they have previously visited while the resource is available.[26] Floral constancy is likely a learned behaviour that increases foraging efficiency during a single trip.[8] Some bees are more selective as far as pollen is concerned. Solitary bee groups show either polylectic or oligolectic foraging behaviour. When visiting only a single species of plant, bees can be said to be monolectic, but behaviours mostly tend to range from narrowly to broadly oligolectic, with the boundaries between them remaining unclear.[27] Oligolectic bees still visit flowers from plant taxa other than those from which pollen are collected for other resources, such as nectar and oils.[12]

Floral choices of bee pollinators play an important role in the sustainability of a plant community. According to food web theory, the more complex

the plant–pollinator interactions are, the less susceptible the plant community is to disturbances or extinction.[28] If one of the interactions fails for some reason, this interaction would likely be taken over by some other pollinator involved in the complex interaction matrix. In contrast, a plant community with a high level of pollinator specialisation would be markedly more vulnerable to any disturbance in its interactions. Plant communities with high diversity would therefore be able to sustain an increased level of bee specialisation, whereas a low diversity plant community would evolve to increase the complexity level of its plant–pollinator interactions. This hypothesis was experimentally verified by increasing plant diversity in a gradient and showing an increase in solitary bee specialisation as plant species richness increased.[29] Oligolectic bees are also more susceptible to changes in their environment and thus to extinction. Because specialist bees have a more restricted foraging range, their effective population size (N_e) and levels of genetic variation are lowered, making these bees and pollination systems vulnerable and in need of protection.[30]

Studying bee-plant interactions

Plant–pollinator interactions have historically been studied through careful and patient observation. This observation usually involves lengthy field-based experiments[2] with plant species in a demarcated area studied for a specific time to see which, if any, animals visit the flowers.[2,18] However, an animal that visits a plant is not necessarily a pollinator of that plant. Even when pollen is transferred to a receptive stigma, genetic incompatibility between pollen and plant may still prevent fertilisation from taking place, as pollen tubes may not germinate, pollen tube growth down the style may be terminated, or pollen may simply be unviable.[31] Fertilisation could be unsuccessful because the pollen being deposited onto the stigma is from a different plant species. Self-incompatibility also prevents fertilisation by pollen from the same plant.[31] Laboratory experiments using captive pollinators can also be conducted, especially for confirmation studies, but these studies do not reflect the pollinators' natural environment.[32]

Field-based observation can be followed by determination of the pollen loads on potential pollinators and the assessment of pollination effectiveness, as measured by the degree of fruiting and seed set through examinations of the individual plants visited.[33] Pollen load determination of any potential pollinator requires capturing the animal in question and the removal of the pollen it carries. The pollen morphology is then carefully analysed (palynology) to identify, or confirm, the plant species from which it originates. Palynology-based identification requires sufficient knowledge of the field and intimate familiarity and expertise with pollen morphological structures, especially of closely related plant species. Furthermore, some form of microscopy is required for visualising the pollen's morphological features used in the identification process, such as scanning electron microscopy or compound light microscopy.[31,34] These technologies require specialised sample preparation methods, skill to prepare and operate instruments, and experience to best obtain comparative morphological features between samples. Additionally, pollen morphological features of different plant species or genera can be extremely similar, especially if they are closely related, thereby requiring a wide palynology knowledge base to accurately distinguish between these samples.[34] Mixed pollen samples from closely related species would therefore require a highly skilled and knowledgeable palynologist, usually an expert familiar with the pollen from the area under investigation. The pollen-carrier must also be identified to make accurate inferences,[2] a function normally performed by different taxonomic specialists in entomology. This process makes plant–pollinator interaction studies time-consuming and highly multi-disciplinary, and requires expertise in the fields of taxonomy, botanical reproduction, palynology, entomology and microscopy.

Pollinator declines

The most well-known bee species is the honeybee, *Apis mellifera* Linnaeus, and most bee-related research has focused on this species.[6,35] Honeybee populations have been reported to be declining in certain areas of the world, such as central Europe, the United States of America and Mexico.[35,36] Although there is as yet no consensus on what may be driving *A. mellifera* population decline, factors such as insecticide use on crops; pests, diseases and predators; a decrease in genetic variation of bee colonies; the effects of climate change; and limitations in the trade of bee colonies may all play a role.[35,36]

In South Africa and some other countries, honeybee numbers are seemingly not declining. This is attributed to beekeeping (apiculture) and the past intentional introduction of numerous alien plant species, which widened the honeybee foraging range.[37] Honeybee colonies in South Africa were seen to be resilient to most diseases. This view was supported when an outbreak of American foulbrood in 2011 did not cause any major colony losses.[38] More recently in 2015, however, another American foulbrood outbreak in South Africa reduced the number of colonies in the Western Cape by 40%.[39]

It is not only honeybee populations that have been declining over time. Researchers in Britain and the Netherlands have found a correlation between declines in native bees and declines in the number of outcrossing plant species dependent on these bees.[40] The native bees in both countries have narrow habitat requirements and produce single broods per year. Honeybee data were specifically excluded, but data for all native species for both countries were included in the analyses. The ultimate cause and direction of the declines could not be determined from the data, but the aforementioned study supports the notion that species reliant on a wider range of interactions within a plant–pollinator system would be more resilient when threatened.[40]

A changing climate, inappropriate land-management and a growing human population have contributed to the reduction of overall biodiversity, including native, wild bee populations across the world.[40] An important determinant of the maintenance of plant–pollinator interactions is the way land is used and managed.[6,41] When agriculture is intensified on a piece of land, bee diversity can decline because of fewer opportunities for them to nest, lower foraging diversity and possible insecticide use on crops.[6] In South Africa's Karoo, all of these factors have been documented to result in a decline of bee and wasp diversity.[4] Game farming started to replace stock farming in this region and land is often overexploited. Tourism opportunities availed from game farming resulted in the introduction of animal species not normally found in the area, and also no period of rest for the land to recover. In areas where large plots of single cultivated plants (monocultures) are found, such as in the wheat fields and wine lands of the Western Cape, very little of the natural vegetation remains. These areas are also likely sprayed with pesticides. In combination, these factors can cause the complete loss of native bee and wasp communities.[4]

Managing pollinators

Native honeybees are currently the only pollinators that are being managed in South Africa. The management of honeybee colonies for pollination purposes has several advantages and disadvantages. One major advantage of using honeybees in pollination management is their generalist foraging habits that make them suitable to be used on many different crop species. Like many other bee species, they are nevertheless unable to pollinate all crops.[41] Additionally, they pack their collected pollen into the corbiculae on their hind legs after moistening it with nectar or honey. This results in limited pollen available for pollination[8] and renders the honeybee a poorer pollinator when compared to other bee species[25]. African honeybees are also aggressive and care needs to be taken when working with them.[41] Their susceptibility to pesticides, diseases and parasites also threatens their commerciality[35] and it is consequently important that pollination management strategies using other native species be explored.

Crop production has increased dramatically over the past decades to meet the demands of growing populations. This growth means that pollinator population sizes are not adequate to deal with the demand. Managed pollinators provide a solution to this problem. It would be best to manage indigenous pollinator populations, such as the honeybee in Africa, to alleviate the problem. In light of some of the inadequacies of honeybees as pollinators, and it being the only group of managed pollinators in South Africa, investigation into the floral choices of native bees could

identify candidates for management in the place of, or in addition to, honeybee populations. It is possible that an indigenous carpenter bee, *Xylocopa scioensis*, could be managed for tomato pollination in South Africa, much like the leafcutter bee, *Megachile rotundata*, is used for alfalfa (lucerne) pollination in North America.[42] Tomatoes require vibratile ('buzz') pollination, something not effectively achieved by honeybees. This indigenous species could potentially avoid the need to import foreign pollinators as has been proposed for the Western region of South Africa, to aid in the pollination of particularly vibratile pollinated crops.[43]

Effects of Bombus *introduction*

The introduction of generalist pollinators, like the honeybee, to provide pollination services to multiple crop species is an economic choice. The honeybee (*A. mellifera*) naturally occurs throughout Africa, Europe and western Asia, but has been introduced to a significant proportion of the rest of the world as a successful pollinator.[22] The European bumblebee, *Bombus terrestris*, is most commonly found throughout Europe but has also recently been introduced into other countries[44,45], and its effect on the environment is well documented. Suggestions have been made that South Africa would benefit from introducing *B. terrestris* and managing their populations for pollination[43] as South Africa does not have any native *Bombus* species.

The impact of the introduction of any alien species into an environment should therefore be carefully considered, as it could be devastating to native ecosystems. This effect was demonstrated with two indigenous subspecies of *A. mellifera* in South Africa: *A. m. capensis* and *A. m. scutellata*. When colonies of *A. m. capensis* were moved across the hybrid zone separating these subspecies in the early 1990s, tens of thousands of *A. m. scutelata* colonies were lost.[46] The loss occurred because of the ability of *A. m. capensis* to infest *A. m. scutelata* colonies and then establish a female clonal lineage that parasitises on *A. m. scutellata* colonies – a trait unique to *A. m. capensis.* These infestations have been shown to only occur when humans transport *A. m. capensis* colonies into *A. m. scutelata*'s native habitat.[47]

Similarly, since its introduction into foreign habitats, *B. terrestris* has had major effects on native plant and bee populations, both positive and negative. It has been shown to increase pollination overall, but decrease efficiency of pollination in native plants, enhance pollination in weeds, and cause displacement of native pollinators.[44] It was discovered, by chance, to have invaded Neuquén Province in Argentina during a survey of floral visitors of shrubs. The bees were thought to have entered Argentina from Chile, as extensive studies of natural and museum populations in Argentina did not provide any historical evidence for the presence of *B. terrestris*.[45] Analysis of the pollen found on the *B. terrestris* individuals showed that they were competing with an indigenous *Bombus* species for food on seven out of the eight host plants. In Japan, *B. terrestris* was introduced to pollinate crops, but then escaped from greenhouses, became naturalised and has had negative consequences on the native bee populations.[48] Resource competition between the introduced and native bumblebees was found in the Japanese study. Introduced species also interfered with the reproduction of both native plants and native bumblebees (by interspecies crosses). Additionally, new parasites were introduced to native populations. Native bumblebee populations have been displaced by *B. terrestris* before,[49] making its invasiveness of great concern. Previously, it has invaded Tasmanian national and urban gardens where it was found foraging on a wide variety of plant types.[44]

Therefore, the use of a foreign pollinator in South Africa should be carefully considered. Native, oligolectic bee species in South Africa would be particularly vulnerable to an introduction of *B. terrestris*, or any other polylectic species, that would be managed for pollination services. So far, permits have not been granted to import *B. terrestris* into South Africa, but in February 2014, Senegal received a shipment of *B. terrestris* colonies from Belgium[50], signifying the first introduction of this species in sub-Saharan Africa[43]. The preceding evidence clearly indicates the possibility that the bees introduced elsewhere in Africa could spread to South Africa.

Harnessing genetic methods to examine floral choice

The high species diversity of both plants and pollinators in South Africa makes the traditional methodology for studying plant–pollinator interactions cumbersome and impractical, particularly in projects encompassing many different species of plants and pollinators. Additionally, the few published works in this area suggests limited expertise within this field worldwide. Another approach is therefore needed to investigate these interactions more efficiently. Genetic methods can prove advantageous in revealing the floral choice patterns of native bees in South Africa. Insect taxonomists across the country have built, and are constantly adding to, large collections of native bees sampled from all over the country. Many of these bees have pollen attached to their bodies that can be used to genetically determine the taxa of plants that they visited in the flight before they were collected. Plant species within the country are also currently being collected, identified and barcoded.[51]

DNA barcoding

DNA barcoding has been successfully used as a diagnostic tool to identify morphologically cryptic species (by comparison to reference libraries) and has highlighted previously unrecognised species, for example various fish and amphipod crustacean species.[52-54] The genetic barcoding of a specimen involves the amplification of a DNA region that has a higher level of interspecific variation and limited intraspecific divergence. Gene regions used in barcoding should also provide a DNA target that can be easily amplified across many taxa using universal primers.[52,55]

The application of DNA barcoding to identify pollen has only recently been developed. In the last 2 years, research publications utilising barcoding in palynology have increased.[56-61] The use of next-generation sequencing (NGS) in identifying mixed pollen samples (pollen metabarcoding) has recently become possible, with a full laboratory protocol and bioinformatics analysis pipeline published.[61]

Pollen as a template for genetic studies

How pollen is collected can impact the success of downstream molecular applications. When sampling pollen from hives, nests, or even honey, sufficient sample quantities are usually available for processing.[57,58,61,62] Conversely, when sampling pollen directly from pollinators, specimens may have only very limited pollen grains captured on their bodies. Small sample quantities may limit and complicate all further laboratory steps. The physical structure of pollen can also be problematic when used as a template. Pollen has an extremely hardy outer wall to protect it from various environmental factors, and this wall could influence DNA extraction and other processes.[63] Different methods are currently used to extract DNA from pollen, but most include a step to disrupt the tough pollen exine.[57-59] A standardised DNA extraction method for pollen barcoding purposes would greatly aid in the comparability among studies.[56]

Using a NGS approach for plant–pollinator interaction studies allows the collection and barcoding of pollen, even if only a few pollen grains are available. Pollinator specimens can therefore be used as pollen sources. When bees from a natural collection are to be used as a pollen source, some factors need to be kept in mind. Bees might have extremely limited quantities of pollen captured on their bodies, and DNA extraction and all subsequent steps should be optimised for use with low starting DNA concentrations in mind. The manner in which the collection has been maintained is also of primary concern. It is well known that bees collect fungi together with pollen[64], but a collection kept in suboptimal conditions would see additional fungal and bacterial growth[65]. Depending on the research question, barcoding gene regions can be selected to amplify more than just plant DNA from pollen samples.

Amplifying DNA from pollen samples

The barcoding principle was first applied to animal groups using the mitochondrial cytochrome c oxidase I gene (*COI*).[52] In plants, the mitochondrial gene variation is not as great between species as in

animals thus making the use of the *COI* barcode region ineffective as a barcode within the Plant Kingdom. Many studies have been done to search for a suite of barcode markers for use in land plants, with varied outcomes and numerous suggestions of genes to target.[66,67] The focus has mainly been on the plastid genome, with ribulose-1,5-biphosphate carboxylase oxygenase (*rbcL*) and maturase K (*matK*) being the most studied genes and, at first glance, the most informative. The Consortium for the Barcode of Life Plant Working Group was established to develop all aspects with regard to plant barcoding. They have suggested the use of *rbcL* and *matK* as the standard barcode for plants[66] after evaluating the success of combinations of coding regions (*matK*, *rbcL*, *rpoB* and *rpoC1*) and non-coding regions (*atpF–atpH*, *trnH–psbA*, and *psbK–psbI*). The internal transcribed spacer (ITS) region of the nuclear genome has been suggested as an additional region to barcode, with several of the plastid genes added to increase identification success.[68]

However, the plastid genome is usually uniparentally inherited; and amplifying plastid DNA could potentially present a problem when evaluating pollen from plants with only maternal plastid inheritance.[69] When plastids are exclusively maternally inherited, the ITS barcode could be invaluable in identifying the pollen parent plant as pollen contains two sperm cells which contain the nuclear genome of the plant.[70,71] Sometimes organellar DNA is biparentally inherited and some plastid leakage from the non-contributing parent can also occur.[69]

PCR and sequencing

Sequencing pollen DNA has initially been done directly from the PCR template by traditional Sanger sequencing. A study on Hawaiian *Hylaeus* bees investigated the pollen composition in the bees' guts to determine their pollen foraging behaviour.[72] ITS barcodes were sequenced for samples that were preserved in 100% ethanol post-collection and the 28S ribosomal RNA region for samples preserved in 70% ethanol. Using Sanger sequencing, most pollen samples could only be identified to one plant species, but in two samples pollen belonging to two species could be identified. For mixed pollen samples, PCR products have been cloned and subsequently Sanger sequenced.[62,73] In this approach, a number of clones are picked and sequenced prior to identification against a reference database. Sequencing clones of pollen found in multiflower honey produced identifications for between 12 and 15 taxa per sample and 38 taxa overall[62], and pollen from honeybee pollen pellets collected from hives produced between 21 and 31 taxa per sample and 52 taxa overall[73]. Pollen identifications were made using the Basic Local Alignment Search Tool (BLAST) available for searching GenBank, which provides a best-hit similarity method of analysis.[74] However, using a cloning approach to obtain single identifiable barcodes from a mixed sample with a number of unknown species, is time consuming and expensive, with no known way to confirm that all the species have been sequenced successfully.

NGS has made it possible to process many samples simultaneously as a result of the parallel nature of the technology. It is hence much more cost-effective to sequence mixed-origin samples on an NGS platform.[75] Each PCR strand is sequenced separately in NGS and this eliminates the need for prior microscopic sorting or cloning of mixed pollen samples. Studies published recently in the pollen barcoding field have combined barcoding with NGS as the preferred sequencing method.[57-61] For example, metabarcoding was used to investigate the floral composition of honey samples in commercial[76] and domestic beekeeper-provided honeys[57]. A larger region of the same barcode as was used in metabarcoding commercial honeys[76] – the chloroplast *trn*L (UAA) intron region – was used to test the efficiency of NGS in identifying the plant origins of airborne pollen[59]. A chloroplast gene was also used as barcode in the study on beekeeper-provided honeys[57], whereas several others[58,60,61] were successful using a nuclear region for pollen identification. Different sequencing platforms have been used for pollen metabarcoding; these platforms are summarised in Table 1 together with the particular study's application in the field and choice of genetic barcode marker.

Table 1: A comparison of the recently published pollen metabarcoding studies, focusing on the application of the study, the barcode region selected and the next-generation sequencing platform used in the study

Pollen metabarcoding application	Barcode region	Next-generation sequencing platform
Aeroallergen monitoring[59]	*trn*L	Ion Torrent PGM
Provenance monitoring[60]	ITS2	Illumina MiSeq
Provenance monitoring[58]	ITS2	Roche 454 GS junior
Provenance monitoring[61]	ITS2	Illumina MiSeq
Food quality and provenance monitoring[57]	*rbcL*	Roche 454 GS FLX
Food quality and provenance monitoring[76]	*trn*L	Roche 454 GS 20

Multiplexed samples need a way to be separated post-sequencing. Adding unique sequence indexes to the sequencing adapters in NGS systems allows this to be done bioinformatically.[61] Various indexing methods have also been successfully used, with dual indexing of PCR products by far the most cost-effective as it allows for a higher degree of multiplexing. Illumina has published a workflow for 16S metagenomics that adds overhang adapters to gene-specific PCR primers, from which dual-indexing can be done directly with the Nextera® XT (Illumina Part #15044223 Rev. B) indexing PCR. This protocol can be adapted for use in any metagenomic application, making it ideal for metabarcoding of pollen.

Bioinformatics

The incorporation of NGS in the barcoding process produces considerable amounts of data. Bioinformatic pipelines catering to the specific metabarcoding needs of pollen analysis are essential to provide reliable identifications using sequence reference databases. Sequence similarity (or best-hit) approaches[74] have long been in use, but suffer from some drawbacks. Heuristic searches on local alignments are performed, and a value is given of the probability that another equally good hit will be found by chance. This value is not comparable to a confidence score and relates only to the local alignment, not the taxonomic assignment of the sequence.[76,77] Other software available for bacterial taxonomy assignments use classifiers, such as the Ribosomal Database Project Classifier[78] and the UTAX command in USEARCH (currently not published; http://www.drive5.com/usearch/manual/utax_algo.html) and these tend to perform better than best-hit approaches. Classifiers rely on the assignment of information in a hierarchical manner to provide taxonomic classifications together with a confidence score. Incorrect assignments can still be made when classifiers are trained on incomplete or incorrect sequence reference databases. Recently, a complete bioinformatics pipeline has been published for ITS2[61], providing much needed guidance to researchers in the field. Standardised bioinformatics methods still need to be developed so that data can be easily analysed across different studies.

The reliable use of barcoding in species identification requires high-quality sequence databases that connect specific species to their DNA barcodes and that hierarchically connect these species taxonomically. These features are particularly important when mixed-species pollen, such as that sampled from a bee's body, is being assigned to its taxonomic origin during analysis. Additionally, a database is required for each barcode region used, with the availability of barcode sequences for these databases dependent on the usage of the DNA region within the taxon under investigation. Most sequence databases are comprised of sequences obtained from publicly available databases, such as GenBank at the National Center for Biotechnology Information (NCBI). This scenario is not ideal, as misidentified entries could be present and often the relevant barcode markers are not available in these public databases, thereby resulting in gaps for the gene region of interest in the barcode reference database. Additionally, incomplete barcoding of plant

species within a region further compounds correct identifications. Pollen samples will subsequently be mismatched to available sequences in the reference database or left unidentified. Some sequences are available in databases that undergo quality checks, such as the ITS2-Database.[79] Plant data contained in the Barcode of Life Database Systems (BOLD; www.boldsystems.org[80]) are all from *rbcL* and *matK*, the two proposed plant plastid barcodes. These sequences are submitted by researchers and must conform to certain standards to be accepted. The available ITS2-Database has also recently been expanded nearly 2.5 times for plants.[61] Curated databases provide higher confidence in the underlying sequence data, whereas sequences in NCBI are often taxonomically misclassified, but sometimes represent the only available entry for a particular species. This limitation could lead to the underestimation of within-species diversity as a result of recent speciation.[81] Bioinformatics methods applied to barcoding sequence data are consequently a crucial part of producing reliable pollen identifications.

Conclusion

Given South Africa's rich flowering plant and bee diversity, the immense economic significance of pollination for agriculture, and the threats of climate change and poor land management on the country's biodiversity, investigations into plant–pollinator relationships are vital. Floral choice in bees gives a good indication of which plants they likely pollinate. Oligolectic bees are more vulnerable to upsets in their relationships with plants. As it has been suggested that the Succulent Karoo Biome in the western part of South Africa contains many oligolectic species, this is a key region of interest for study. Should bumblebees be introduced to this area, as suggested previously, much of the bee biodiversity of South Africa could be at stake. The identification of pollen origins is important in understanding the floral choices of bees. Many advances have been made in recent years in molecular pollen identification. DNA metabarcoding can provide accurate taxonomic identifications of pollen origins when compared to comprehensive sequence databases of carefully selected barcode gene regions. However, the lack of barcoding information for the bulk of the South African flora is a major stumbling block still to be overcome. Pollen from both honeybees and their honey, and solitary bees, has successfully been identified using this technique. DNA metabarcoding should prove instrumental in the exploration of floral choice in South African bees.

Acknowledgements

We thank the National Research Foundation of South Africa (grant 95649) and the Belgian Directorate-General for Development Cooperation, partim GTI for funding the research. A.G. also thanks the South African Department of Science and Technology/National Research Foundation Professional Development Programme (99781), the Biotechnology Platform at the Agricultural Research Council, and the University of KwaZulu-Natal for support. We also thank the anonymous reviewers who made extremely useful comments which lead to the improvement of our manuscript.

Authors' contributions

Z.H.S., C.E. and S.W.M. were the project leaders; A.G., Z.H.S., C.E. and S.W.M. were responsible for project design; A.G., Z.H.S., C.E. and S.W.M. made conceptual contributions; and Z.H.S., C.E. and S.W.M. supervised A.G. in her PhD studies. A.G. wrote and revised the manuscript, and Z.H.S., C.E. and S.W.M. provided critical reviews of the manuscript.

References

1. Mittermeier RA, Turner WR, Larsen FW, Brooks TM, Gascon C. Global biodiversity conservation: The critical role of hotspots. In: Zachos FE, Habel JC, editors. Biodiversity hotspots. Heidelberg: Springer; 2011. p. 3–22. https://doi.org/10.1007/978-3-642-20992-5_1
2. Gess SK, Gess FW. A comparative overview of flower visiting by non-*Apis* bees in the semi-arid to arid areas of southern Africa. J Kans Entomol Soc. 2004;77(4):602–618. http://dx.doi.org/10.2317/E7.1
3. South African Department of Environmental Affairs. Land. In: 2nd South Africa Environment Outlook: A report on the state of the environment. Pretoria: Department of Environmental Affairs; 2012. p. 87–105.
4. Gess SK, Gess FW. Wasps and bees in southern Africa. Pretoria: South African National Biodiversity Institute; 2014.
5. Losey JE, Vaughan M. The economic value of ecological services provided by insects. BioScience. 2006;56(4):311–323. http://dx.doi.org/10.1641/0006-3568(2006)56[311:TEVOES]2.0.CO;2
6. Klein A-M, Vaissiere BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, et al. Importance of pollinators in changing landscapes for world crops. Proc R Soc B Biol Sci. 2007;274(1608):303–313. http://dx.doi.org/10.1098/rspb.2006.3721
7. Kearns CA, Inouye DW, Waser NM. Endangered mutualisms: The conservation of plant–pollinator interactions. Annu Rev Ecol Syst. 1998;29:83–112. http://dx.doi.org/10.1146/annurev.ecolsys.29.1.83
8. Michener CD. The bees of the world. 2nd ed. Baltimore: Johns Hopkins University Press; 2000.
9. Johnson SD. The pollination niche and its role in the diversification and maintenance of the southern African flora. Philos Trans R Soc B Biol Sci. 2010;365(1539):499–516. http://dx.doi.org/10.1098/rstb.2009.0243
10. Minckley RL, Cane JH, Kervin L. Origins and ecological consequences of pollen specialization among desert bees. Proc R Soc Lond B Biol Sci. 2000;267(1440):265–271. http://dx.doi.org/10.1098/rspb.2000.0996
11. Padyšáková E, Bartoš M, Tropek R, Janeček Š. Generalization versus specialization in pollination systems: Visitors, thieves, and pollinators of *Hypoestes aristata* (Acanthaceae). PLoS ONE. 2013;8(4), e59299, 8 pages. http://dx.doi.org/10.1371/journal.pone.0059299
12. Bosch J, Martín González AM, Rodrigo A, Navarro D. Plant–pollinator networks: Adding the pollinator's perspective. Ecol Lett. 2009;12(5):409–419. http://dx.doi.org/10.1111/j.1461-0248.2009.01296.x
13. Hetherington-Rauth MC, Ramírez SR. Evolutionary trends and specialization in the euglossine bee–pollinated orchid genus *Gongora*. Ann Mo Bot Gard. 2015;100(4):271–299. http://dx.doi.org/10.3417/2014035
14. Johnson SD. Pollination, adaptation and speciation models in the Cape flora of South Africa. Taxon. 1996;45(1):59–66. http://dx.doi.org/10.2307/1222585
15. De Merxem DG, Borremans B, De Jäger ML, Johnson T, Jooste M, Ros P, et al. The importance of flower visitors not predicted by floral syndromes. S Afr J Bot. 2009;75(4):660–667. http://dx.doi.org/10.1016/j.sajb.2009.08.002
16. Johnson SD, Steiner KE. Generalization versus specialization in plant pollination systems. Trends Ecol Evol. 2000;15(4):140–143. http://dx.doi.org/10.1016/S0169-5347(99)01811-X
17. Bawa KS. Plant–pollinator interactions in tropical rain forests. Annu Rev Ecol Syst. 1990;21:399–422. http://dx.doi.org/10.1146/annurev.es.21.110190.002151
18. Kwapong PK, Danquah POA, Asare AT. Insect floral visitors of cowpea (*Vigna unguiculata* L.). Ann Biol Res. 2013;4(4):12–18.
19. Kuhlmann M. Patterns of diversity, endemism and distribution of bees (Insecta: Hymenoptera: Anthophila) in southern Africa. S Afr J Bot. 2009;75(4):726–738. http://dx.doi.org/10.1016/j.sajb.2009.06.016
20. Eardley CD. Diversity and endemism of southern African bees. Plant Protection News. 1989 Dec;18:1–2.
21. Welsford MR, Johnson SD. Solitary and social bees as pollinators of *Wahlenbergia* (Campanulaceae): Single-visit effectiveness, overnight sheltering and responses to flower colour. Arthropod-Plant Interact. 2012;6(1):1–14. http://dx.doi.org/10.1007/s11829-011-9149-0
22. Goulson D. Effects of introduced bees of native ecosystems. Annu Rev Ecol Syst. 2003;34:1–26. http://dx.doi.org/10.1146/annurev.ecolsys.34.011802.132355
23. Ginsberg HS. Historical development of bee foraging patterns in central New York State. Psyche (Stuttg). 1981;88(3–4):337–346. http://dx.doi.org/10.1155/1981/29040
24. Thorp RW. The collection of pollen by bees. Plant Syst Evol. 2000;222:211–223. http://dx.doi.org/10.1007/BF00984103
25. Westerkamp C. Honeybees are poor pollinators – why? Plant Syst Evol. 1991;177(1–2):71–75. http://dx.doi.org/10.1007/BF00937827

26. Wilson P, Stine M. Floral constancy in bumble bees: Handling efficiency or perceptual conditioning? Oecologia. 1996;106(4):493–499. http://dx.doi.org/10.1007/BF00329707

27. Linsley EG. The ecology of solitary bees. Vol. 27. Berkeley, CA: University of California; 1958.

28. Melián CJ, Bascompte J. Food web structure and habitat loss. Ecol Lett. 2002;5(1):37–46. http://dx.doi.org/10.1046/j.1461-0248.2002.00280.x

29. Ebeling A, Klein A-M, Tscharntke T. Plant–flower visitor interaction webs: Temporal stability and pollinator specialization increases along an experimental plant diversity gradient. Basic Appl Ecol. 2011;12(4):300–309. http://dx.doi.org/10.1016/j.baae.2011.04.005

30. Packer L, Zayed A, Grixti JC, Ruz L, Owen RE, Vivallo F, et al. Conservation genetics of potentially endangered mutualisms: Reduced levels of genetic variation in specialist versus generalist bees. Conserv Biol. 2005;19(1):195–202. http://dx.doi.org/10.1111/j.1523-1739.2005.00601.x

31. Dafni A. Pollination ecology: A practical approach. Oxford: Oxford University Press; 1992.

32. Harder LD. Choice of individual flowers by bumble bees: Interaction of morphology, time and energy. Behaviour. 1988;104(1/2):60–77. http://dx.doi.org/10.1163/156853988X00601

33. Klein A-M, Steffan-Dewenter I, Tscharntke T. Fruit set of highland coffee increases with the diversity of pollinating bees. Proc R Soc B Biol Sci. 2003;270(1518):955–961. http://dx.doi.org/10.1098/rspb.2002.2306

34. Rahl M. Microscopic identification and purity determination of pollen grains. In: Jones MG, Lympany P, editors. Allergy methods and protocols. Totowa, NJ: Humana Press; 2008. https://doi.org/10.1007/978-1-59745-366-0_22

35. VanEngelsdorp D, Meixner MD. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. J Invertebr Pathol. 2010;103:S80–S95. http://dx.doi.org/10.1016/j.jip.2009.06.011

36. Goulson D, Nicholls E, Botías C, Rotheray EL. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science. 2015;347(6229), Art. #1255957. http://dx.doi.org/10.1126/science.1255957

37. Jaffé R, Dietemann V, Allsopp MH, Costa C, Crewe RM, Dall'Olio R, et al. Estimating the density of honeybee colonies across their natural range to fill the gap in pollinator decline censuses. Conserv Biol. 2010;24(2):583–593. http://dx.doi.org/10.1111/j.1523-1739.2009.01331.x

38. Human H, Pirk CWW, Crewe RM, Dietemann V. The honeybee disease American foulbrood – An African perspective. Afr Entomol. 2011;19(3):551–557. http://dx.doi.org/10.4001/003.019.0301

39. Kings S. Honeybee crisis catches SA off-guard. Mail & Guardian. 2015 April 17 [cited 2016 Jun 01]. Available from: http://mg.co.za/article/2015-04-16-honeybee-crisis-catches-sa-off-guard

40. Biesmeijer JC, Roberts SPM, Reemer M, Ohlemüller R, Edwards M, Peeters T, et al. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. Science. 2006;313:351–353. http://dx.doi.org/10.1126/science.1129551

41. Kearns CA, Inouye DW. Pollinators, flowering plants, and conservation biology. BioScience. 1997;47(5):297–307. http://dx.doi.org/10.2307/1313191

42. Pitts-Singer TL, Cane JH. The alfalfa leafcutting bee, *Megachile rotundata* : The world's most intensively managed solitary bee. Annu Rev Entomol. 2011;56(1):221–237. http://dx.doi.org/10.1146/annurev-ento-120709-144836

43. Rodger J, Donaldson J, Eardley C, Pauw A, Johnson S. SA should put bumblebees to flight. Farmer's Weekly. 2013 [updated 2014 Jul 09; cited 2016 Mar 12]. Available from: http://www.farmersweekly.co.za/article.aspx?id=40333&h=SA-should-put-bumblebees-to-flight

44. Hingston AB, Marsden-Smedley J, Driscoll DA, Corbett S, Fenton J, Anderson R, et al. Extent of invasion of Tasmanian native vegetation by the exotic bumblebee *Bombus terrestris* (Apoidea: Apidae). Austral Ecol. 2002;27(2):162–172. http://dx.doi.org/10.1046/j.1442-9993.2002.01179.x

45. Torretta JP, Medan D, Abrahamovich AH. First record of the invasive bumblebee *Bombus terrestris* (L.)(Hymenoptera, Apidae) in Argentina. Trans Am Entomol Soc. 2006;132:285–289. http://dx.doi.org/10.3157/0002-8320(2006)132[285:FROTIB]2.0.CO;2

46. Byatt MA, Chapman NC, Latty T, Oldroyd BP. The genetic consequences of the anthropogenic movement of social bees. Insectes Sociaux. 2016;63(1):15–24. https://doi.org/10.1007/s00040-015-0441-3

47. Beekman M, Allsopp MH, Wossler TC, Oldroyd BP. Factors affecting the dynamics of the honeybee (*Apis mellifera*) hybrid zone of South Africa. Heredity. 2008;100(1):13–18. http://dx.doi.org/10.1038/sj.hdy.6801058

48. Goka K. Introduction to the special feature for ecological risk assessment of introduced bumblebees: Status of the European bumblebee, *Bombus terrestris*, in Japan as a beneficial pollinator and an invasive alien species. Appl Entomol Zool. 2010;45(1):1–6. http://dx.doi.org/10.1303/aez.2010.1

49. Inoue MN, Yokoyama J, Washitani I. Displacement of Japanese native bumblebees by the recently introduced *Bombus terrestris* (L.) (Hymenoptera: Apidae). J Insect Conserv. 2008;12(2):135–146. http://dx.doi.org/10.1007/s10841-007-9071-z

50. Siegmund H. Brussels airlines cargo welcomes 16 queen bees aboard. CargoForwarder Global. 2014 March 03 [cited 2016 Mar 12]. Available from: http://www.cargoforwarder.eu/2014/03/03/brussels-airlines-cargo-welcomes-16-queen-bees-aboard/

51. Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, et al. DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci USA. 2008;105(8):2923–2928. https://doi.org/10.1073/pnas.0709936105

52. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc R Soc B Biol Sci. 2003;270(1512):313–321. http://dx.doi.org/10.1098/rspb.2002.2218

53. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. DNA barcoding Australia's fish species. Philos Trans R Soc B Biol Sci. 2005;360(1462):1847–1857. http://dx.doi.org/10.1098/rstb.2005.1716

54. Witt JDS, Threloff DL, Hebert PDN. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: Implications for desert spring conservation. Mol Ecol. 2006;15(10):3073–3082. http://dx.doi.org/10.1111/j.1365-294X.2006.02999.x

55. Moritz C, Cicero C. DNA barcoding: Promise and pitfalls. PLoS Biol. 2004;2(10), e354, 3 pages. http://dx.doi.org/10.1371/journal.pbio.0020354

56. Bell K, De Vere N, Keller A, Richardson R, Gous A, Burgess K, et al. Pollen DNA barcoding: Current applications and future prospects. Genome. 2016;59:629–640. http://dx.doi.org/10.1139/gen-2015-0200

57. Hawkins J, De Vere N, Griffith A, Ford CR, Allainguillaume J, Hegarty MJ, et al. Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. PLoS ONE. 2015;10(8), e0134735, 20 pages. http://dx.doi.org/10.1371/journal.pone.0134735

58. Keller A, Danner N, Grimmer G, Ankenbrand M, Von der Ohe K, Von der Ohe W, et al. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol. 2015;17(2):558–566. http://dx.doi.org/10.1111/plb.12251

59. Kraaijeveld K, De Weger LA, Ventayol García M, Buermans H, Frank J, Hiemstra PS, et al. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol Ecol Resour. 2015;15(1):8–16. http://dx.doi.org/10.1111/1755-0998.12288

60. Richardson RT, Lin C-H, Quijia JO, Riusech NS, Goodell K, Johnson RM. Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. Appl Plant Sci. 2015;3(11), 1500043, 9 pages. http://dx.doi.org/10.3732/apps.1500043

61. Sickel W, Ankenbrand MJ, Grimmer G, Holzschuh A, Härtel S, Lanzen J, et al. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecol. 2015;15(1), Art. #20, 9 pages. http://dx.doi.org/10.1186/s12898-015-0051-y

62. Bruni I, Galimberti A, Caridi L, Scaccabarozzi D, De Mattia F, Casiraghi M, et al. A DNA barcoding approach to identify plant species in multiflower honey. Food Chem. 2015;170:308–315. http://dx.doi.org/10.1016/j.foodchem.2014.08.060

63. Edlund AF. Pollen and stigma structure and function: The role of diversity in pollination. Plant Cell. 2004;16:S84–S97. http://dx.doi.org/10.1105/tpc.015800

64. Eltz T, Brühl CA, Görke C. Collection of mold (*Rhizopus* sp.) spores in lieu of pollen by the stingless bee *Trigona collina*. Insectes Sociaux. 2002;49(1):28–30. http://dx.doi.org/10.1007/s00040-002-8274-2

65. Merritt J. Mold: Prevention of growth in museum collections. Conserve O Gram. 2007;3/4:1–5.

66. Consortium for the Barcode of Life Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci USA. 2009;106(31):12794–12797. http://dx.doi.org/10.1073/pnas.0905845106

67. Kress WJ, Erickson DL. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE. 2007;2(6), e508, 10 pages. http://dx.doi.org/10.1371/journal.pone.0000508

68. Yao H, Song J, Liu C, Luo K, Han J, Li Y, et al. Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE. 2010;5(10), e13102, 9 pages. http://dx.doi.org/10.1371/journal.pone.0013102

69. Corriveau JL, Goff LJ, Coleman AW. Plastid DNA is not detectable in the male gametes and pollen tubes of an angiosperm (*Antirrhinum majus*) that is maternal for plastid inheritance. Curr Genet. 1990;17(5):439–444. http://dx.doi.org/10.1007/BF00334525

70. McCue AD, Cresti M, Feijo JA, Slotkin RK. Cytoplasmic connection of sperm cells to the pollen vegetative cell nucleus: potential roles of the male germ unit revisited. J Exp Bot. 2011;62(5):1621–1631. http://dx.doi.org/10.1093/jxb/err032

71. Nagata N, Saito C, Sakai A, Kuroiwa H, Kuroiwa T. The selective increase or decrease of organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance. Planta. 1999;209(1):53–65. http://dx.doi.org/10.1007/s004250050606

72. Wilson EE, Sidhu CS, LeVan KE, Holway DA. Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis. Mol Ecol. 2010;19(21):4823–4829. http://dx.doi.org/10.1111/j.1365-294X.2010.04849.x

73. Galimberti A, De Mattia F, Bruni I, Scaccabarozzi D, Sandionigi A, Barbuto M, et al. A DNA barcoding approach to characterize pollen collected by honeybees. PLoS ONE. 2014;9(10), e109363, 13 pages. https://doi.org/10.1371/journal.pone.0109363

74. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2

75. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012, Art. #251364, 11 pages. http://dx.doi.org/10.1155/2012/251364

76. Valentini A, Miquel C, Taberlet P. DNA barcoding for honey biodiversity. Diversity. 2010;2(4):610–617. http://dx.doi.org/10.3390/d2040610

77. Munch K, Boomsma W, Huelsenbeck J, Willerslev E, Nielsen R. Statistical assignment of DNA sequences using Bayesian phylogenetics. Syst Biol. 2008;57(5):750–757. http://dx.doi.org/10.1080/10635150802422316

78. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73(16):5261–5267. http://dx.doi.org/10.1128/AEM.00062-07

79. Keller A, Schleicher T, Schultz J, Müller T, Dandekar T, Wolf M. 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. Gene. 2009;430(1–2):50–57. http://dx.doi.org/10.1016/j.gene.2008.10.012

80. Ratnasingham S, Hebert PD. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Mol Ecol Notes. 2007;7(3):355–364. http://dx.doi.org/10.1111/j.1471-8286.2007.01678.x

81. Sandionigi A, Galimberti A, Labra M, Ferri E, Panunzi E, De Mattia F, et al. Analytical approaches for DNA barcoding data – How to find a way for plants? Plant Biosyst. 2012;146(4):805–813. http://dx.doi.org/10.1080/11263504.2012.740084

Resilience processes in sexually abused adolescent girls: A scoping review of the literature

AUTHORS:
Sadiyya Haffejee[1]
Linda Theron[2]

AFFILIATIONS:
[1]Optentia Research Focus Area, North-West University, Vanderbijlpark, South Africa
[2]Department of Education Psychology, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Sadiyya Haffejee

EMAIL:
sadiyya.haffejee@gmail.com

DATES:
Received: 21 Oct. 2016
Revised: 02 Mar. 2017
Accepted: 23 May 2017

KEYWORDS:
resilience-enabling factors; child sexual abuse; teenage girls; supportive ecologies; positive adaptation

HOW TO CITE:
Haffejee S, Theron L. Resilience processes in sexually abused adolescent girls: A scoping review of the literature. S Afr J Sci. 2017;113(9/10), Art. #2016-0318, 9 pages. http://dx.doi.org/10.17159/sajs.2017/20160318

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
Optentia Research Focus, North-West University; Networks of Change and Well-being Project, International Development Research Centre

Childhood sexual abuse is often associated with a number of deleterious psychological and behavioural outcomes for survivors. However, some research suggests that this impact is variable and that some survivors adapt positively. An ability to adapt positively to adversity, under any circumstances, has been termed resilience. Drawing on a socio-ecological understanding of resilience, the aim of this scoping review was to comprehensively map existing empirical studies on resilience processes in sexually abused adolescent girls and to summarise emerging resilience-enabling factors. We also considered the implications of the findings for practice and research. A total of 11 articles met the criteria for inclusion in the review. Findings from these studies suggest that internal factors (meaning making, optimistic future orientation, agency and mastery) and contextual factors (supportive family, social and educational environments) function interdependently to enable resilience in sexually abused adolescent girls. Practitioners should leverage these complementary and interdependent resilience-enabling mechanisms by encouraging greater involvement of girls in the planning of interventions and by assisting girls in developing meaningful narratives about their abuse experiences. Interventions should also encourage greater involvement from supportive structures, while challenging social and cultural norms that inhibit resilience. Resilience researchers should be cognisant of the paucity of research focusing on resilience processes in sexually abused adolescent girls as well as the absence of innovative, participatory methods of data collection.

Significance:
- The review adds to a body of literature on resilience processes with implications for resilience researchers.
- The findings have implications for a range of practitioners (psychologists, social workers, teachers etc.) who work with sexually abused girls.

Introduction

In August 2016 the hashtag #1in3 was trending in South Africa. This hashtag represents the pervasiveness of sexual abuse in South Africa (i.e. one in three South Africans has experienced sexual abuse). Following a silent protest at a political meeting, women from diverse backgrounds across South Africa took to social media to share experiences of sexual assault, bringing to the fore South Africa's violent rape statistics. One Twitter user tweeted the following[1]:

> *Actually let me just get this straight. In my immediate family – we are 3/3. My mom. My sister. Me. #1in3*

Notwithstanding definitional issues as to what constitutes child sexual abuse (CSA), there is agreement on the pervasiveness of CSA in society, with data suggesting an average worldwide rate of 18% to 20%.[2] South African CSA data are more alarming: Artz et al.[3] found that one in three young people in South Africa have experienced some form of sexual abuse, suggesting that at least 784 967 young people, as of 2016, had been victims of sexual abuse by the time they reached 17 years.

Reading through a number of tweets bearing testimony to the high sexual violence levels in our society, one tweet stands out as most pertinent to this review[4]:

> *That little girl that was raped at 15 is doing well, I look myself in the mirror and smile, I've worked so hard to be where I am today. #1in3*

This ability to adapt positively to adversity has been the focus of resilience research for the past four decades.[5] Resilience researchers address the question of why and how some people adapt positively in spite of significant adversity. This review begins with a similar question, but is distinguished by its specific focus on adolescent girls who have experienced CSA.

Extant literature[2,6,7] suggests that CSA is associated with a number of adverse psychological and behavioural consequences for survivors throughout their lifetime. Depression, substance abuse, eating disorders, dissociation, somatisation, anxiety disorders, post-traumatic stress disorders, psychotic and schizophrenic disorders, and suicidal ideation are among the mental health sequelae linked to CSA.[8-12] CSA has also been correlated with interpersonal, relationship and sexual difficulties.[6,9,13,14] Despite this literature and as evidenced in the tweet above, the impact of CSA is variable; research shows that not all individuals who have experienced CSA develop post-traumatic stress disorder or other psychiatric and behavioural problems.[9,15-17] Some research suggests that less than one fifth of CSA survivors show symptoms of serious psychopathology.[18,19]

Given the high prevalence of sexual violence globally, augmented insights into the resilience processes that buffer the impacts of CSA are particularly relevant. Greater insight into resilience processes potentiates an enhanced understanding of how best to support youth who are at risk for negative developmental outcomes.[20] A cursory

review of scholarly research suggests a profusion of studies on sexual violence, its determinants and its impact for survivors.[2,3,6-9] Although Phasha[21] and Resnick et al.[12] point out that much of this enquiry has overlooked issues pertaining to resilience, internationally there is a growing body of primary studies that focus on resilience in the aftermath of CSA[16]. This development is encouraging given the contribution that these insights can make to specialised clinical services and theoretically supported interventions with CSA survivors. This burgeoning interest has also resulted in a limited number of reviews.[16,17,22] None of these reviews has, however, focused specifically on resilience processes in adolescent girls who have survived CSA. The reviews by Domhardt et al.[17] and Marriot et al.[16] included men and women of all ages, whilst Lekganya's[22] systematic review focused on studies with adult women. Masten and Wright[23] posit that a developmental perspective is necessary when understanding risk and resilience processes; exposure to risks varies with development, as does the individual's processing of the event. Similarly, Supkoff et al.[24] suggest that resilience is a developmental process. An individual may be resilient in one developmental period, but not in another. For example, DuMont et al.[25] found that resilience decreased by about 18% between adolescence and young adulthood, giving credence to the dynamic nature of resilience processes. Thus, an individual demonstrating resilience in adulthood may not necessarily confirm or explain resilience processes during adolescence.

The need to focus only on adolescent girls' experiences stems from research that talks to the role that gender plays in both increasing girls' vulnerability to sexual assault and the buffering effect it may have in enabling resilience. Research shows that while both girls and boys are at risk of sexual abuse, girls are more vulnerable.[26] Patriarchal constructions of masculinity, gender roles assigned to women, social inequality, cultural and parenting practices, and inadequate legal systems increase women's and girls' vulnerability to abuse and also impact their help-seeking behaviour and recovery processes.[27-29] Conversely, Jefferis[30] adds that while girls may be at higher risk, cultural and contextual factors may also enable strength uniquely for girls. Hirani et al.[31] assert that gendered understandings of resilience are necessary as gender roles interact with social and environmental factors to differentially influence how men and women experience and respond to adversity.

Our primary aims in this scoping review are thus to assess the breadth of empirical studies that focus on the resilience of adolescent girls who experienced CSA and to distil the resilience-enabling processes that are common across these studies. Consequently, we consider how findings relating to what enables resilience may be used to inform interventions with adolescent survivors of CSA. We also consider how findings can be used to inform research practices (e.g. choice of research methodologies) within this field and population.

Understanding resilience processes

Although the prevailing critiques of resilience studies often cite the lack of definitional consistency, resilience broadly refers to a process of positive adaptation in the context of severe adversity.[23,32,33] This definition infers that resilience is an interactive process, with two critical aspects present: significant adversity and positive adaptation following exposure to adversity.[34,35] Rutter[36] urges attention to understanding which resources and processes are crucial for adaptation in a given developmental period and/or particular context. This emphasis relates to resilience being a dynamic process. Put differently, resilience is not a single or static individual quality; instead resilience is a process that varies relative to the type of adversity, contextual variables and developmental phase.[35,37]

The reference to 'dynamic' processes also points to the current ecological systems (also called 'social ecological' – see Ungar[38]) understanding of resilience. A social ecological understanding of resilience frames the review that we report in this article. This understanding sees individuals as embedded in dynamic ecological systems that enable positive adaptation.[30,38] Resilience is, thus, defined as the mutually constructive relationship between an individual and his/her ecological system.[38] This reciprocity tasks the individual's social environment to make resilience-enabling resources available, while the individual is simultaneously tasked with moving towards these resources and using them effectively.[39]

Four decades of research has identified universally occurring protective mechanisms that appear to inform resilience.[23,40] Masten[41] suggests that these mechanisms reflect adaptive systems that have been influenced by biological and cultural evolution. They go beyond the individual into other social and cultural systems. This underscores the importance of the environment reiterated by other resilience researchers[42-43] and references the social ecological context to which Ungar[44] speaks.

Masten[41(p.6)] refers to these mechanisms as 'the short list'. This list (itemised below) draws on qualities of the individual and the social ecology.[23] As summarised by Masten[41,45] and Masten and Wright[23], fundamental adaptive systems include:

- attachment relationships and social support that provide capable, responsive caregiving (in adolescence, this support may take the form of close peer relationships);
- problem-solving and the presence of adequate cognitive abilities that allow adequate information processing (resilience does not require exceptional intelligence, but rather the ability to determine what is happening and what to do[23]);
- self-regulation skills that allow for the employment of effective emotional and behavioural regulation strategies;
- agency, mastery and self-efficacy, including a positive sense of self, the presence of self-confidence, the motivation to do something differently to succeed, and a sense of control of the environment;
- meaning making (the ability to find meaning or purpose in all experiences), and hope for a better future, justice or better afterlife;
- the influence of culture, traditions and religion as captured in the presence of protective factors such as faith and traditional and cultural belief systems that assign meaning to experiences. Affiliation with religious communities may also provide support and assistance.

Although the above recur in accounts of resilience, Rutter[36] cautions that the optimal clinical facilitation of the above protective processes requires an understanding of how the type of adversity, context and/or developmental stage alters their impact and meaningfulness. Thus, to address what is known and not known about how the above resilience processes inform the resilience of adolescent girls who have experienced CSA, we opted for a scoping review of the current literature. We used the above shortlist to structure the findings.

Methodology

The scoping review

Scoping reviews have gained increasing popularity as a way of synthesising research findings[46,47] and are broadly a way of determining the research available on a specific topic[48]. Specifically, they are a form of knowledge synthesis that involves mapping existing literature to get a sense of the breadth and depth of a particular research area.[47,49-51] Scoping reviews are typically undertaken to examine the extent, range and nature of the research activity; determine the feasibility of undertaking a full systematic review; summarise and disseminate research findings; and identify gaps in the existing literature.[49] As detailed in the introduction, our review matches the aforementioned given the intention to synthesise what is currently known about the resilience processes of adolescent girls who have experienced CSA. This type of scoping review does not stop at the point of summarisation and dissemination of findings, but goes further to try and draw intervention- and research-related conclusions from the existing literature.[49]

In keeping with the iterative nature of scoping reviews[49], our scoping review took place in two phases. Whereas scoping reviews generally begin with a broader, less focused research question which may then be refined, ours began with too narrow a research question. Initially, the focus was only on qualitative and mixed-methods studies of resilience processes in sexually abused adolescent girls. This first phase was implemented between August 2015 and December 2015. However, after recognising the relative dearth of qualitative studies, we expanded the

study scope to include a wider range of study designs, leading to Phase Two (February 2016 to August 2016). Phase Two included qualitative, quantitative and mixed-methods studies of resilience processes in sexually abused adolescent girls. In both phases, the same search terms and search engines were used. Keywords entered and included in the title and abstract were: (1) resilience OR resilient OR resiliency OR functional outcomes OR positive adjustment OR protective factors AND (2) sexual violence OR rape OR child sexual abuse OR sexual molestation OR sexual assault AND (3) adolescent girls OR girls OR young women. While this inverted process was time-consuming, it allowed for a more comprehensive sense of the existing literature.

Arksey and O'Malley's[49] iterative five-stage model guided this review (Table 1). Where appropriate, the recommendations made by Levac et al.[47] were included. The sections below offer a description of these stages as incorporated in this review.

Identifying the research question

This scoping activity sought to explore factors promoting resilience in adolescent girls who had experienced CSA. The research question thus focuses on adolescents as opposed to adults or young children, on girls as opposed to women or men, and on sexual abuse and not on other forms of abuse.

Identifying relevant studies

Scoping studies requires transparency and rigour, which are ensured through the comprehensiveness of the data search.[49] To achieve this comprehensiveness, we searched a number of different sources: electronic databases, reference lists, university repositories, an online social networking research site, and organisational reports. To ensure access to the correct search engine, we consulted a librarian at North-West University, who provided suggestions regarding keyword combinations and the most comprehensive database to access.

Levac et al.[47] suggest that while comprehensiveness and rigour are important, researchers must be cognisant of practicalities such as time, personnel, and budget constraints. Given the reality of these constraints in our review, we established the following inclusion criteria:

- The study was reported in English.
- The study was a primary study. Theoretical overviews, literature syntheses and studies describing therapeutic interventions were excluded.
- The study population comprised only adolescent girls between the ages of 13 and 19 or where the mean age of the sample was below 20 years. The rationale for limiting the review to adolescent girls' resilience processes has been discussed above. Essentially we concur that gender adds a layer of complexity to resilience processes[30] and as a dynamic process, resilience is subject to change over an individual's lifespan[25,52,53].
- The study focused on resilience processes in girls who had experienced one or more incidents of CSA. We assumed that the authors of each study had adhered to the expectation that resilience studies include only participants who experience a negative event (such as CSA) as adverse and who maintain/regain functional outcomes despite the experience of adversity.[54] Studies covering broader incidents of maltreatment, such as neglect or physical and verbal abuse, were excluded. Some researchers[55] suggest that it is important to separate different types of childhood adversity, as the association between the exposure and outcome may vary, with different maltreatment types likely to predict different outcomes.
- The study referred to resilience processes or protective factors (individual strengths *as well as* social-ecological supports found in families, communities, cultural heritage and so forth). As indicated above, resilience was conceptualised as a social-ecological process informed by the universally occurring processes suggested by Masten and Wright[23]. Studies that addressed one or more of these cardinal processes were included.
- The study was conducted between 1995 and 2016.

Search strategy

Using the keywords and inclusion criteria referred to above, we identified relevant studies via OneSearch, which is a comprehensive electronic search platform that accesses multiple resources, including online resources, journals and books, in one search. Databases in OneSearch that we selected were ERIC, JSTOR, MEDLINE, PsychINFO, PsychARTICLES, ScienceDirect, Academic Search Premier and CINHAL. We also ran separate searches in Scopus, ScienceDirect, JSTOR and Web of Science.

In addition to searching electronic databases, we conducted a search of South African university repositories, including those of the Universities of Stellenbosch, Pretoria, the Western Cape, the Witwatersrand and Johannesburg. Through this process, we identified six South African dissertations. One dissertation met the inclusion criteria.

Table 1: Arksey and O'Malley's[49] five-stage model

Step	Description of step
Identify the research question	Identify the research question to guide the next stages. Important facets or aspects of the question must be clearly defined, as they have implications for the search strategy used.
Identify relevant studies	Identify relevant studies by determining search terms, sources, and language. Levac et al.[47] assert that comprehensiveness and breadth are important in this process. Sources may include electronic databases, university repositories, reference lists and organisational reports.
Select studies	Identify *post hoc* inclusion and exclusion criteria, which can be determined by increased familiarity with the literature.
Chart data	Develop a mechanism, such as a data charting form, to sort through material according to key issues and themes. A descriptive analytical method or narrative review may be used to obtain process-oriented or contextual information.
Collate, report and summarise data	Provide a literature overview, for which an analytic framework or thematic construction is needed. Often a numerical analysis of the extent and nature of studies and a thematic analysis are presented.
Consult with stakeholders (optional)	Invite key stakeholders to suggest additional resources/references and provide additional insights that may not be accessible through the literature review.

We then conducted a manual search of selected articles and literature syntheses that focused on resilience processes and were recommended by resilience researchers or emerged from the electronic search.[16,17,22,30] Through this process, we identified five studies; saturation was reached when no new studies were identified.[49]

We conducted a broad Google search on sexual violence and/or resilience and identified organisations locally and internationally that conducted related research. We then looked through the organisations' outputs, identifying three studies from one organisation, none of which met the inclusion criteria.

Lastly, we accessed an online social networking research site, ResearchGate, to search for specific authors who had previously published on CSA and resilience. In this manner, we accessed two conference papers and two articles. All of these were excluded because of their inclusion of mixed gender samples.

Study selection

The search generated 985 studies. Titles and abstracts were first screened against the inclusion criteria. In this process, 832 were excluded, and 153 were selected for full-text screening, of which 11 were included in this review. Articles were typically excluded because they did not report girl-specific findings. Figure 1 provides a description of the selection process.

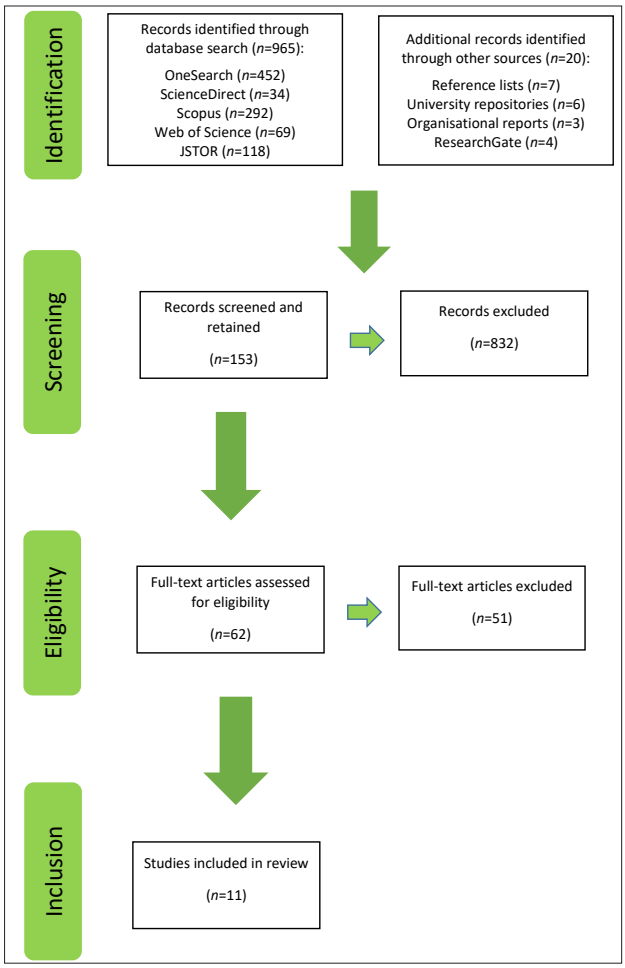


Figure 1: Description of the selection process of studies for review.

The eligibility of only a small number of studies is not surprising, given the focused nature of this review and the nature of the enquiry. As reported by Daigneault et al.[56], a limited number of studies focus on resilience processes in adolescents who have experienced sexual abuse. Other more inclusive reviews of resilience processes in individuals who experienced CSA yielded more studies. For example, Marriot et al.'s[16] narrative literature review of resilient outcomes for people (of all ages and genders) who had experienced CSA included 50 studies, and Domhardt et al.'s[17] systematic review of resilience processes in survivors (men, women, all age groups) yielded 37 studies.

Charting the data

On completion of the selection process, we adopted a descriptive-analytic approach to chart the data, which involved applying a standardised analytical framework to all included studies.[49] We recorded the following information from each of the studies:

- Author(s), year of publication and location (country) of the study
- Source type
- Study population
- Study design/methodology
- Aim of the study
- Results/findings

Appendix 1 in the supplementary material offers a summary of the studies included.

Collating, summarising and reporting the results

We conducted a numerical analysis of the extent and nature of the studies included. Of the 11 studies, 4 adopted a research qualitative design, 1 made use of mixed methods, and 6 were quantitative. Only one study employed a participatory approach. Studies were somewhat evenly distributed among low- and middle-income (South Africa, Sierra Leone and Uganda) and high-income (USA and Canada) countries, with five in the former[18,21,57-59] and six in the latter[56,60-64].

In addition to the above, we used the shortlist summarised by Masten and Wright[23] and Masten[41] to deductively analyse the contents of the included studies. This shortlist (described above) formed the *a priori* codes with which we coded the findings sections of the studies. We summarised these findings into two broad domains for ease of understanding: individual factors and resilience-enabling ecologies. In this way, the scope of available literature, as well as key themes across the 11 studies, is summatively presented. Nevertheless, the reader is cautioned that these resilience-enabling factors function interdependently.

Individual factors

Individual factors encompass the ways in which girls contributed to their resilience trajectories by drawing on personal strengths and capabilities. Factors included their capacity to make meaning of and or reappraise the abusive event, displays of agency and self-motivation, self-efficacy, future-oriented beliefs and self-reliance.

Self-reliance and individual resources were perhaps necessary for the participants in some of the reviewed studies. For example, two studies included participants residing in a foster home/care facility[57,62], two included girls who were referred to child protection services, indicating multiple risks in the family[56,61], and two included African girls in post-war contexts[58,59], again suggesting multiple risk factors and disrupted family and community structures, with traditionally supportive environments (such as families) absent.

Opportunities for meaning making, self-regulation and self-efficacy

Included studies suggested that when girls were able to find meaning, they had less negative appraisals and thought about the abuse experiences differently, thereby minimising their significance and consequently adjusting in more positive ways.[18,21,56,64]

Collings's[18] findings associated resilience processes with appraisal of the sexual abuse as either positive or negative, with negative appraisals suggestive of greater psychopathology. In Daigneault et al.'s[56] study, the ability to make meaning and to integrate memory and affect was a

protective factor linked to disclosure; thus, girls who chose to disclose abuse experiences to a greater number of people had higher scores on meaning making and integration. Their study[56] highlights the significance of disclosure for positive reframing and meaning making in the context of supportive relationships; girls were able to form and maintain relationships with people to whom they had disclosed. At follow-up with a sample of participants in this study[56], symptom relief and improved resilience were found to be associated with participants' ability to feel and think about the incident differently[61]. Also apparent were higher scores in the domains of self-esteem.

Phasha[21], exploring educational resilience in the aftermath of CSA, found that the young women in her study adapted positively by making sense of their experiences of sexual violence as temporary or learning experiences devoid of self-blame. The absence of self-blaming attitudes meant that the participants were not hampered by thoughts of guilt and helplessness:

> *I accepted that it's in the past. I mustn't put it in my mind because it will condemn my mind.... I know I have a future even though I did go through that bad stuff …* [21]

Similarly, participants in Himelein and McElrath's[63] study also made use of positive reframing and sought to view the incident as a learning experience or vehicle for change. This belief resulted in new feelings of strength and maturity. As in Phasha's[21] study, this attempt to find meaning in the adversity conveyed a sense of hope and optimism for the participants: 'Once I was able to put it aside … it made a huge difference.'[63]

Self-regulation was apparent in Archer's[57] study; for the 16-year-old participant in her study, placement at a residential facility following sexual abuse contributed to greater independence, expressions of agency, and regulation of her behaviour. Through careful observation and interaction with the care system, she understood what actions she needed to take to manoeuvre successfully and made active choices in controlling her emotions and avoiding trouble. This ability to self-regulate translated into greater ownership of her behaviour and a less blaming attitude.

Optimistic future perspectives

Future orientated beliefs and the presence of goals and aspirations are associated with a positive perception of the world and efforts to adapt in contexts of adversity.[65] This factor appeared to play a role in studies of sexually abused girls as well. Edmond et al.[62] found the presence of optimism towards the future characteristic of the resilient girls in their study.

Likewise, participants in Phasha's[21] study were able to envision a future free from abuse and viewed the future positively rather than with hopelessness: 'My future is very important to me.'[21]

Archer[57] also found evidence of future planning and positive thoughts in her study. Optimistic future planning was evident in both short- and long-term plans. In the quote below, we see the ability to plan for the future, a demonstration of agency, as well as behavioural self-regulation, which highlights the interconnectedness of the various resilience-promoting mechanisms: 'I have a number of dreams for my future … I must learn to control my temper better as well if I want to become an air hostess …'[57].

Agency and motivational mastery

Jefferis[30] refers to 'agentic women' when referencing women's and girls' displays of resolve and decision-making in the face of adversity. This display of volition was apparent in the participants in Denov and MacLure's[58] study, in which girls demonstrated resilience by exercising individual agency through personal choices to actively participate or subtly resist oppressive circumstances. This suggests that, in dealing with sexual violence, girls are not merely subservient victims, but rather exhibit a capacity for independent thought and action.[58] In much the same way, Archer[57] showed that agency extended towards accessing social and legal resources, even if such consequences might be to the detriment of maternal support or material circumstances. Gilligan et al.[66] refer to this display of assertion as a way of children retaining a sense of themselves as independent actors and actively seeking out resources that may be of use to them.

Resilience-enabling ecologies

An ecologically grounded understanding of resilience shifts the onus of responsibility for resilience away from the child and instead focuses on the environment to facilitate and make resources available to the child.[44] Resources take the form of supportive family environments, peer relationships, social and cultural communities and educational systems.

Attachment relationships

Luthar[67] maintains that relationships are central to resilience processes. The significance of relationships is corroborated in a qualitative synthesis of resilience mechanisms in girls: Jefferis[30] found that, out of the 38 included studies, every study referred to the presence of a constructive relational context that included positive relationships with other people, spiritual beings or other animals. This finding is further evidenced in this review, with supportive attachment systems emerging as a resilience enabler in 7 of the 11 studies reviewed.

Aspelmeier et al.[60] caution that attachment does not completely ameliorate negative CSA outcomes, although it does offer some protection. Findings from Spaccerelli and Kim[64] corroborate this statement, with results showing that the presence of supportive parents was related to average or above average social competence and appeared to play an important role in maintaining school performance, activities and peer relationships after the abuse. Here parental support was associated with the ability to appraise the incident in a less negative way. In Archer[57], maternal support was linked to the adolescent's sense of identity. Stark et al.[59] also found that maternal support was a protective factor against stigma associated with sexual abuse: 'My mother tells me not to worry. She comforts me.'[59]

These supportive relationships were not limited to parents, but also extended to peers and other sources of support. Denov and MacLure[58] found that the girls in their study often relied on other girls in similar situations or older women who had experienced similar difficulties and who were thus able to share information and guide them. This sense of solidarity eased the difficulties they encountered, and the older women created safe spaces and a sense of nurturance:

> *We ... started sharing our stories* [of rape] *… I felt much better after this because I thought that I was the only one to have this happen … the older women helped me.*[58]

The significance of peer relationships during adolescence is widely accepted in developmental literature. The studies included echo the importance of these relationships, with exposure to less negative peer behaviour associated with resilience.[62] Stark et al.[59] reported that girls drew on close friends to help mitigate stigma. Aspelmeier et al.[60] found that higher levels of secure peer attachment/relationships, determined by perceived quality of communication, were related to lower levels of self-related trauma symptoms among CSA survivors. In both Archer's[57] and Phasha's[21] studies, this circle of support was broadened, with participants seeking support and assistance from teachers, friends, social workers and police officers.

The role of supportive environments did not appear limited to offering nurturing and encouragement; in Stark et al.[59], parents and other community members, such as officials and police officers, played a role in negotiating some form of justice. While this role may have implied that the participants might be 'powerless'[59(p.222)] in negotiating their own fate in certain circumstances, in other circumstances, it pointed to girls' agency in understanding how legal systems worked and how they could afford protection. For example, the participant in Archer's[57] study understood early on that reporting abuse would mean removal from her home. While separation from the non-offending parent was difficult, she also understood that such separation offered protection. Both scenarios highlight the possibilities that supportive ecologies hold in these contexts.

Cultural and religious traditions

Two of the South African studies included found evidence of the protective impact of religious beliefs on sexually abused adolescent girls. Phasha[21] found that African girls often made meaning of CSA by understanding it as God's will and not their personal responsibility. Phasha linked this belief, which acted as a source of strength and a way to acceptance, to African child-rearing practices being deeply entrenched in religious beliefs and understandings.

In Archer's[57] study, the belief in a higher being allowed for greater meaning making and acceptance; the participant's belief in God and regular church and spiritual group attendance assisted her in forgiving those whom she perceived as complicit in the abuse.

Spiritual attributions or meaning making, which included beliefs in 'white magic' and 'spirits', were also noted in a Canadian sample[61]; however, this finding was specific to the Haitian youth in that sample. The researchers viewed this finding cautiously, given the test instrument's limited cultural sensitivity.

Social, community and educational systems

The importance of school contexts and education was highlighted in three studies.[21,57,62] Educational success, engagement in school systems, and certainty regarding educational futures were noted. Phasha[21] explained the desire for educational success and commitment to schooling for the girls in her study in terms of racial and ethnic identity. Phasha[21] maintains that self-knowledge and self-regard – processes that partially stemmed from a positive racial identity and a strong sense of ethnic identity – allowed African participants to overcome adversity. These beliefs grew, in part, because of apartheid policies of discrimination and, simultaneously, conveyed the importance of education as a vehicle to a better future. In Archer's[57] study, getting involved in educational and recreational activities provided a sense of belonging and success, bolstering feelings of confidence and self-esteem.

Edmond et al.[62] showed that girls with a resilient trajectory appeared to skip school less frequently, were less likely to get into fights at school, and appeared to have a greater sense of certainty regarding their educational futures than did the comparison sample of symptomatic girls. Approximately 88% of participants with resilient trajectories indicated that they were very sure of their high school plans, with three out of four planning on going to college.[62] Findings in this study were reinforced in that participants also scored higher on the future orientation scale, suggesting a more positive view of their futures, and thereby associating education with a better future.

Discussion

Our aim in conducting this scoping review was to examine and understand what was known globally about resilience processes in sexually abused adolescent girls as well as how this insight may be used to inform clinical practice, intervention and research. A review of the 11 studies included suggests a correlation with protective factors identified in the broader literature on resilience processes in youth. Furthermore, we note the presence of internal and contextual factors that function interdependently to enable resilience in sexually abused adolescent girls. What is novel about these findings is that they are the first to offer an overview of the resilience processes that support adolescent girls from Global South and North contexts to adjust well following CSA.

Taking heed of Levac et al.'s[47] counsel to reviewers to consider the implications of the results and what they may mean for 'research, policy and practice', we consider what the findings from this review imply for resilience researchers and practitioners working with sexually abused girls.

Implications for practice

Findings from the review highlight possible points at which practitioners (e.g. psychologists, social workers, child care workers and health workers) may act in order to leverage resilience-enabling practice. These actions include assisting girls in the development of meaningful narratives about their abuse experiences; greater consultation with, and inclusion of, affected girls in planning of interventions; challenging of social and cultural norms; and eliciting and encouraging greater involvement from supportive structures.

The findings suggest that adolescent girls themselves are resourceful and display agency in responding to sexual abuse. Cognitive reappraisals and positive reframing of the incident impacted feelings of self-efficacy and allowed some to minimise the impact CSA had on them and their futures; these girls were, thus, able to view the future optimistically. Opportunities for intervention in fostering resilience are thus apparent: without minimising the experience, girls can be assisted to develop a meaningful narrative of the abuse and to envisage a future.[56] Collings[18] suggests that cognitive appraisals can be altered; thus, employing cognitive change strategies holds promise in interventions with this group.

Denov and MacLure[58] also draw on their findings of agency demonstrated by participants to recommend that interventions acknowledge the volition of adolescent girls and intervene in ways whereby the voices and perspectives of youth are elicited and incorporated in intervention programmes. Interventions should provide opportunities for girls' 'demonstrable capacity of thought and action' to be acknowledged.[58(p.82)]

We found strong evidence of the role of attachment relationships in enabling resilience trajectories, beginning with supportive parents and extending to enabling community members, highlighting the importance of these structures. In contrast, Stark et al.[59] reported that while social networks appeared to be vital in offering support and promoting psychosocial recovery in the aftermath of abuse, such networks did not consistently respond positively to the needs of the adolescent girls. In some instances, girls were stigmatised, blamed or socially excluded by family and community members because of the sexual abuse; in these instances, social ecologies impeded recovery processes.[59] Jefferis[30] also notes that social, cultural and community norms around gender roles – which value women's and girls' submissiveness and prioritise roles of caretaking – sometimes act to the detriment of resilience processes for girls and women. Stark et al.[59], speaking to their findings on the role of stigma and shame, recommend that interventions address destructive social norms that perpetuate stigma. Thus, interventions aimed at adolescent girls who have experienced CSA should take cognisance of the potential and risks that social ecologies, such as families and communities, present.

Implications for future research

What is apparent from this review is the paucity of research focusing on resilience processes in sexually abused adolescent girls. After a thorough search strategy using multiple databases, only 11 exclusive studies were identified. To date, studies on resilience processes in youth have usually incorporated different forms of adversity, seldom focusing on only one risk factor. Furthermore, resilience studies with children and youth often include both male and female participants. As discussed earlier, gender adds a layer of complexity to resilience processes; gender roles, societal norms and contextual factors impact how men and women experience and respond to stressors.[67] Where resilience processes were examined specifically in relation to sexual abuse, participants were often adult women and not adolescent girls. This finding suggests a need for further explorations of resilience processes specific to sexually abused adolescent girls. Limited understanding of resilience processes in this group may result in interventions that are not gender and/or developmentally sensitive, lack coherence for adolescent girls, and may be ineffective.

Although our review revealed a complementary mix of both qualitative and quantitative studies, there was an absence of child-centred participatory approaches. Only one of the included studies[58] engaged girls as co-researchers. This decision was motivated by their belief that the inclusion of girls as co-researchers would enhance the richness of the discussion and thus the quality of data. Additionally, for the researchers[58], participatory methods also held the potential of involving the girls in an educational, empowering and purposeful activity. While there is value in all empirical studies on CSA and resilience processes, regardless

of methodology, we add to a growing call for the inclusion of youth in research about them.[62,68] Alderson[68] asserts that youth are key sources of information about their own experiences and inclusivity results in more insightful research. Participatory methods position youth as knowledge producers. Engaging youth as knowledge producers in research has been successfully implemented in projects focused on finding solutions to social challenges, like HIV, sexual vulnerability and teenage pregnancy.[69] Moreover, participatory methods are youth friendly, allowing for greater accessibility and the redress of power imbalances between the researcher and the participants.[70] Such methods have been recommended for use in studies on both gender violence and youth resilience. Mitchell[71] suggests that studies focused on gender violence require methods that allow for engagement and disengagement of participants, so that participants themselves can structure the pace. Participatory approaches allow for this control. Similarly, Theron[72] suggests that qualitative approaches that embrace creative, innovative data collection techniques result in more comprehensive understandings of youth resilience processes.

What the discussion above has reinforced is that CSA survivors are not passive victims acting without agency; they are capable of reflecting on and negotiating change and solutions to the difficulties they encounter.[73] Participatory methods in particular hold the potential for enhancing this process; a participatory approach creates avenues for youth voices to be heard, encourages displays of agency and provides a platform for empowerment.[74]

Given the possibilities that participatory approaches present in terms of richer, more in-depth understandings generated by engaged youth themselves and their suitability for use specifically with sexually abused girls in terms of potentiating agency, it appears that researchers need to consider these approaches when exploring resilience processes.

Conclusion

This scoping review closely mirrored the first five steps outlined by Arksey and O'Malley[49], and while constrained by the limited number of studies available, it provided valuable insight into the resilience processes of sexually abused adolescent girls. The omission of the optional final step (consultation with key stakeholders to provide additional information and insight) described by Arksey and O'Malley[49] presents a limitation. It is possible that such a dialogue could have provided further knowledge regarding enabling processes and added insight into how this knowledge could be disseminated and translated into practice. Notwithstanding this possibility, the coherence of these findings with the shortlist summarised by Masten and Wright[23] is compelling and, following Rutter[36], adds developmental- and risk-specific knowledge to the body of literature on resilience processes. The greater contribution of the findings is in the possibilities they offer for practitioners and researchers to exact change in adolescent girls' recovery processes and enhance the field of resilience studies in youth.

Acknowledgements

We gratefully acknowledge the PhD bursary provided to S.H. by Optentia Research Focus, North-West University as well as the Networks of Change and Well-being Project (funded by the International Development Research Centre).

Authors' contributions

Both authors conceptualised the article; S.H. wrote up the method, reviewed the literature and drafted the initial themes; L.T. assisted in refining the themes; and both authors contributed to the discussion.

References

1. Hastie N [Nina Hastie@THATninahastie]. Actually let me just get this straight. In my immediate family – we are 3/3. My mom. My sister. Me. #1in3 [Twitter]. 2016 Aug 07 [cited 2016 Aug 10]. Available from: https://twitter.com/thatninahastie
2. Collin-Vezina D, Daigneault I, Hebert M. Lessons learned from child sexual abuse research: Prevalence, outcomes, and preventive strategies. Child Adolesc Psychiatry Ment Health. 2013;7, Art. #22, 9 pages. http://dx.doi.org/10.1186/1753-2000-7-22
3. Artz L, Burton P, Ward C, Leoschut L, Phyfer J, Lloyd S, et al. Sexual victimisation of children in South Africa. Technical report. Zurich: UBS Optimus Foundation; 2016.
4. Matomane N [Nolundi Matomane@NolundiMatomane]. That little girl that was raped at 15 is doing well, I look myself in the mirror and smile, I've worked so hard to be where I am today. #1in3 [Twitter]. 2016 Aug 07 [cited 2016 Aug 10]. Available from: https://twitter.com/nolundimatomane
5. Masten AS. Resilience in developing systems: Progress and promise as the fourth wave rises. Dev Psychopathol. 2007;19:921–930. http://dx.doi.org/10.1017/S0954579407000442
6. Maniglio R. The impact of child sexual abuse on health: A systematic review of reviews. Clin Psychol Rev. 2009;29(7):647–657. http://dx.doi.org/10.1016/j.cpr.2009.08.003
7. Hillberg T, Hamilton-Giachitsis C, Dixon L. Review of meta-analyses on the association between child sexual abuse and adult mental health difficulties: A systematic approach. Trauma Violence Abuse. 2011;12(1):38–49. http://dx.doi.org/10.1177/1524838010386812
8. Cashmore J, Shackel R. The long term effects of child sexual abuse. Melbourne: Australian Institute of Family Studies; 2013. CFCA PAPER NO. 11 2013. http://cdn.basw.co.uk/upload/basw_103914-1.pdf
9. Collishaw S, Pickles A, Messer J, Rutter M, Shearer C, Maughan B. Resilience to adult psychopathology following childhood maltreatment: Evidence from a community sample. Child Abuse Neglect. 2007;31(3):211–229. http://dx.doi.org/10.1016/j.chiabu.2007.02.004
10. Finkelhor D. The prevention of childhood sexual abuse. Future Child. 2009;19(2):169–194. http://dx.doi.org/10.1353/foc.0.0035
11. Kilpatrick DG, Acierno R. Mental health needs of crime victims: Epidemiology and outcomes. J Trauma Stress. 2003;16:119–132. http://dx.doi.org/10.1023/A:1022891005388
12. Resnick HS, Guille C, McCauley JL, Kilpatrick DG. Rape and other sexual assault. In: Southwick SM, Litz BM, Charney D, Friedman MJ, editors. Resilience and mental health: Challenges across a lifespan. New York: Cambridge University Press; 2011. p. 218–237. https://doi.org/10.1017/cbo9780511994791.017
13. Lalor K, McElvaney R. Child sexual abuse, links to later sexual exploitation/high risk sexual behaviour and prevention/treatment programmes. Trauma Violence Abuse. 2010;(11):159–177. http://dx.doi.org/10.1177/1524838010378299
14. Miron LR, Orcutt HK. Pathways from childhood abuse to prospective revictimisation: Depression, sex to reduce negative affect, and forecasted sexual behaviour. Child Abuse Neglect. 2014;38(11):1848–1859. https://doi.org/10.1016/j.chiabu.2014.10.004
15. McElharan M, Briscoe-Smith A, Khaylis A, Westrup D, Hayward C, Gore-Felton C. A conceptual model of post traumatic growth among children and adolescents in the aftermath of sexual abuse. Couns Psychol Q. 2012;25(1):73–82. http://dx.doi.org/10.1080/09515070.2012.665225
16. Marriot C, Hamilton-Giachritsis C, Harrop C. Factors promoting resilience following childhood sexual abuse: A structured, narrative review of the literature. Child Abuse Rev. 2014;23:17–34. http://dx.doi.org/10.1002/car.2258
17. Domhardt M, Munzer A, Fegert JM, Lutz G. Resilience in survivors of child sexual abuse: A systematic review of the literature. Trauma Violence Abuse. 2015;(16)4:476–493. http://dx.doi.org/10.1177/1524838014557288
18. Collings S. Surviving child sexual abuse. The social work practitioner-researcher. 2003;15(1):97–110.
19. Kendall-Tacket KA, Williams LM, Finkelhor D. Impact of sexual abuse on children. A review and synthesis of recent empirical studies. In: Bull, R, editor. Children and the law: The essential readings. Malden, MA: Blackwell Publishers, 1993. Psychological Bulletin, 113(1):164–180. http://dx.doi.org/10.1037/0033-2909.113.1.164
20. Luthar SS, Ciccheti D, Becker B. The construct of resilience: A critical evaluation and guidelines for future work. Child Dev. 2000;71(3):543–562. https://doi.org/10.1111/1467-8624.00164
21. Phasha NT. Educational resilience among African survivors of child sexual abuse in South Africa. J Black Stud. 2010;40(6):1234–1253. http://dx.doi.org/10.1177/0021934708327693

22. Lekganya I. A systematic review of the psychological factors associated with resilience among survivors of sexual abuse [master's thesis]. Cape Town: University of the Western Cape, 2015.

23. Masten A, Wright M. Resilience over the lifespan: Developmental perspectives on resistance, recovery and transformation. In: Reich JW, Zautra AJ, Hall JS, editors. Handbook of adult resilience. New York: The Guilford Press, 2010. p. 213–237.

24. Supkoff LM, Puig J, Sroufe LA. Situating resilience in developmental context. In: Ungar M, editor. The social ecology of resilience: A handbook of theory and practice. Springer: New York, 2012. p. 127–142. http://dx.doi.org/10.1007/978-1-4614-0586-3_12

25. Dumont KA, Widom CS, Czaja SJ. Predictors of resilience in abused and neglected children grown up: The role of individual and neighbourhood characteristics. Child Abuse Neglect. 2007;31:255–274. http://dx.doi.org/10.1016/j.chiabu.2005.11.015

26. Dartnell E, Jewkes R. Sexual violence against women: The scope of the problem. Best Pract Res Clin Obstet Gynaecol. 2013;27:3–13. http://dx.doi.org/10.1016/j.bobgyn.2012.08.002

27. Lalor K. Child sexual abuse in sub-Saharan Africa: Child protection implications for development policy makers and practitioners. Development Research Briefings no. 3. Dublin: Centre for Development Studies, University College Dublin; 2005. Available from: http://arrow.dit.ie/cgi/viewcontent.cgi?article=1001&context=aaschsslrep

28. Mathews S, Loots L, Sikweyiya Y, Jewkes R. Sexual abuse. In: Van Niekerk A, Suffla S, Seedat M, editors. Crime violence and injury in South Africa: 21st century solution for child safety. Johannesburg: Psychological Society of South Africa; 2012. p. 1–7.

29. Seedat M, Van Niekerk A, Jewkes R, Suffla S, Ratele K. Violence and injuries in South Africa: Prioritising an agenda for prevention. Lancet. 2009;374(9694):1011–1022. http://dx.doi.org/10.1016/s0140-6736(09)60948-x

30. Jefferis T. Resilient black South African girls in contexts of adversity: A participatory visual study [PhD thesis]. Vanderbijlpark: North West University; 2016.

31. Hirani S, Lasiuk G, Hegadohen K. The intersection of gender and resilience. J Psychiatr Ment Health Nurs. 2016;23(6–7):455–467. http://dx.doi.org/10.1111/jpm.12313

32. Masten AS. Ordinary magic: Resilience processes in development. Am Psychol. 2001;56(3):227–238. http://dx.doi.org/10.1037/0003-066X.56.3.227

33. Luthar SS, Cicchetti D. The construct of resilience: Implications for interventions and social policies. Dev Psychopathol. 2000;12:857–885. http://dx.doi.org/10.1017/S0954579400004156

34. Wright MO, Masten AS. Pathways to resilience in context. In: Theron LC, Liebenberg L, Ungar M, editors. Youth resilience and culture: Commonalities and complexities. New York: Springer; 2015. p. 3–22. http://dx.doi.org/10.1007/978-94-017-9415-2_1

35. Rutter M. Implications of resilience concepts for scientific understanding. Ann NY Acad Sci. 2006;1094:1–12. http://dx.doi.org/10.1196/annals.1376.002

36. Rutter M. Annual research review: Resilience – Clinical implications. J Child Psychol Psychiatr. 2013:54(4):474–487. http://dx.doi.10.1111/j.1469-7610.2012.02615.x

37. Rutter M. Psychosocial resilience and protective mechanisms. Am J Orthopsychiatry. 1987;57(3):316–331.http://dx.doi.org/10.1111/j.1939-0025.1987.tb03541.x

38. Ungar M. The social ecology of resilience: Addressing contextual and cultural ambiguity of a nascent construct. Am J Orthopsychiatry. 2011;81(1):1–17. http://dx.doi.org/10.1111/j.1939-0025.2010.01067.x

39. Hall Am, Theron L. How school ecologies facilitate resilience among adolescents with intellectual disabilities: Guidelines for teachers. S Afr J Educ. 2016;36(2):1–13. http://dx.doi.org/10.15700/saje.v36n2a1154

40. Wright MO, Masten AS, Narayan AJ. Resilience processes in development: Four waves of research on positive adaptation in the context of adversity. In: Goldstein S, Brooks RB, editors. Handbook of resilience in children. New York: Springer; 2013. p. 15–37. http://dx.doi.org/10.1007/978-1-4614-3661-4_2

41. Masten AS. Promoting resilience in development: A general system for frameworks of care. In: Flynn R, Dudding PM, Barber JG, editors. Promoting resilience in child welfare. Ottawa: University of Ottawa Press; 2006. p. 3–17.

42. Theron LC, Theron A. Positive adjustment to poverty: How family communities encourage resilience in traditional African contexts. Cult Psychol. 2013;19(3):391–413. http://dx.doi.org/10.1177/1354067X13489318

43. Panter-Brick C. Culture and resilience: Next steps for theory and practice. In: Theron LC, Liebenberg L, Ungar M, editors. Youth resilience and culture – Commonalities and complexities. Dordrecht: Springer; 2015. p. 233–244. http://dx.doi.org/10.1007/978-94-017-9415-2_17

44. Ungar M. Resilience across cultures. Br J Soc Work. 2008;38(2):218–235. http://dx.doi.org/10.1093/bjsw/bcl343

45. Masten AS. Resilience in development: Early childhood as a window of opportunity. Presented at: ECHD Symposium; 2015 April 24; Albuquerque, New Mexico, USA.

46. Pham MT, Rajic A, Greig JD, Sargeant JM, Papadopolous A, McEwn SA. A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. Res Syn Methods. 2014;5(4):371–385. https://doi.org/10.1002/jrsm.1123

47. Levac D, Colquhoun H, O'Brien KK. Scoping studies: Advancing the methodology. Implement Sci. 2010;5(1):69–78. https://doi.org/10.1186/1748-5908-5-69

48. Woznowski-Vu A, Da Costa C, Turgeon-Provost F, Dagenais K, Roy-Mathie B, Aggban M, et al. Factors affecting length of stay in adult outpatient physical rehabilitation: A scoping review of the literature. Physiother Can. 2015;67(4):329–340. http://dx.doi.org/10.3138/ptc.2014-75

49. Arskey H, O'Malley L. Scoping studies: Towards a methodological framework. Int J Soc Res Method. 2005;8(1):19–32. https://doi.org/10.1080/1364557032000119616

50. Armstrong R, Hall BJ, Doyle J, Waters E. Cochrane update: Scoping the scope of a Cochrane review. J Public Health. 2011;33(1):147–150. http://dx.doi.org/10.1093/pubmed/fdr015

51. Peters M, Godfrey CM, Mcinerney P, Baldini Soares C. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc. 2015;13:141–146. http://dx.doi.org/10.1097/XEB.0000000000000050

52. Egeland B, Carlson E, Sroufe LA. Resilience as a process. Dev Psychopathol. 1993;5(4):517–528. http://dx.doi.org/10.1017/S0954579400006131

53. Ciccheti D, Rogosch FA. The role of self-organisation in the promotion of resilience in maltreated children. Dev Psychopathol. 1997;9(4):797–815. https://doi.org/10.1017/s0954579497001442

54. Van Rensburg A, Theron L, Rothman S. A review of quantitative studies of South African youth resilience: Some gaps. S Afr J Sci. 2015;111(7–8), Art. #2014-0164, 9 pages. http://dx.doi.org/10.17159/SAJS.2015/20140164

55. Afifi TO, Macmillan HL. Resilience following child maltreatment: A review of protective factors. Can J Psychiatry. 2011;56(5):266–272.

56. Daigneault I, Tourigny M, Cyr M. Description of trauma and resilience in sexually abused adolescents. J Trauma Prac. 2004;3(2):23–47. https://doi.org/10.1300/j189v03n02_02

57. Archer E. Exploring the phenomenon of resilience with a child survivor of abuse [master's thesis]. Pretoria: University of Pretoria; 2005.

58. Denov M, MacLure R. Engaging the voices of girls in the aftermath of Sierra Leone's Conflict: Experiences and perspectives in a culture of violence. Anthropologica. 2006;48(1):73–85. http://dx.doi.org/10.2307/25605298

59. Stark L, Landis D, Thomson B, Potts A. Navigating support, resilience, and care: Exploring the impact of informal social networks on the rehabilitation and care of young female survivors of sexual violence in northern Uganda. Peace Confl. 2016;22(3):217–225. http://dx.doi.org/10.1037/pac0000162

60. Aspelmeier JE, Elliot AN, Smith CH. Childhood sexual abuse, attachment, and trauma symptoms in college females: The moderating role of attachment. Child Abuse Neglect. 2007;31:549–566. http://dx.doi.org/10.1016/j.chiabu.2006.12.002

61. Daigneault I, Cyr M, Tourigny M. Exploration of recovery trajectories in sexually abused adolescents. J Aggres Maltreat Trauma. 2007;14(1–2):165–185. https://doi.org/10.1300/j146v14n01_09

62. Edmond T, Auslander W, Elze D, Bowland S. Signs of resilience in sexually abused adolescent girls in the foster care system. J Child Sex Abuse. 2006;15(1):1–28. http://dx.doi.org/10.1300/J070v15n01_01

63. Himelein MJ, McElrath JV. Resilient child sexual abuse survivors: Cognitive coping and illusion. Child Abuse Neglect. 1996;20(8):747–758. http://dx.doi.org/10.1016/0145-2134(96)00062-2

64. Spaccarelli S, Kim S. Resilience criteria and factors associated with resilience in sexually abused girls. Child Abuse Neglect. 1995;19(9):1171–1182. http://dx.doi.org/10.1016/0145-2134(95)00077-L

65. Malindi MJ. Swimming upstream in the midst of adversity: Exploring resilience-enablers among street children. J Soc Sci. 2014;39(3):265–274.

66. Gilligan,R, De Castro EP, Vanistendael S, Warburton J. Learning from children exposed to sexual abuse and sexual exploitation: Synthesis report of the Bamboo project study on child resilience. Geneva: Oak Foundation; 2004.

67. Luthar SS. Resilience in development: A synthesis of research across five decades. In: Cichetti D, Cohen DJ, editors. Development psychopathology. Vol 3: Risk, disorder and adaptation. New York: Wiley; 2006. p. 739–795.

68. Alderson P. Children as researchers: Participation rights and research methods. In: Christensen P, James A, editors. Research with children: Perspectives and practices. New York: Routledge; 2000. p. 276–291.

69. Stuart J. Youth as knowledge producers: Towards changing gendered patterns in rural school with participatory arts-based approaches to HIV and AIDS. Agenda. 2011;24(84):53–65. http://dx.doi.org/10.1080/10130950.2010.9676309

70. Didkowsky N, Ungar M, Liebenberg L. Using visual methods to capture embedded processes of resilience for youth across cultures and contexts. J Can Acad Child Adolesc Psychiatry. 2010;19(1):2–18.

71. Mitchell C. Doing visual research. London: Sage; 2011.

72. Theron LC. Resilience research with South African youth: Caveats and ethical complexities. S Afr J Psychol. 2012;42(3):333–345. http://dx.doi.org/10.1177/008124631204200305

73. Buthelezi T, Mitchell C, Moletsane R, De Lange N, Taylor M, Stuart J. Youth voices about sex and AIDS: Implications for life skills education through the 'Learning Together' project in KwaZulu-Natal, South Africa. Int J Inclusive Educ. 2007;11(4):445–459. http://dx.doi.org/10.1080/13603110701391410

74. Mitchell C, De Lange N, Stuart J, Moletsane R, Buthelezi T. Children's provocative images of stigma, vulnerability and violence in the age of AIDS: Re-visualisations of childhood. In: De Lange N, Mitchell C, Stuart J, editors. Putting people in the picture. Rotterdam: Sense Publishers; 2007. p. 59–71.

Potential of interval partial least square regression in estimating leaf area index

AUTHORS:
Zolo Kiala[1]
John Odindi[1]
Onisimo Mutanga[1]

AFFILIATION:
[1]School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

CORRESPONDENCE TO:
Zolo Kiala

EMAIL:
serkial1@yahoo.fr

DATES:
Received: 08 Sep. 2016
Revised: 18 Jan. 2017
Accepted: 09 May 2017

KEYWORDS:
hyperspectral data; iPLSR; modelling; partial least square regression; tropical grassland

HOW TO CITE:
Kiala Z, Odindi J, Mutanga O. Potential of interval partial least square regression in estimating leaf area index. S Afr J Sci. 2017;113(9/10), Art. #2016-0277, 9 pages. http://dx.doi.org/10.17159/sajs.2017/20160277

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

Leaf area index (LAI) is a critical parameter in determining vegetation status and health. In tropical grasslands, reliable determination of LAI, useful in determining above ground biomass, provides a basis for rangeland management, conservation and restoration. In this study, interval partial least square regression (iPLSR) in forward mode was compared to partial least square regression (PLSR) to estimate LAI from in-situ canopy hyperspectral data on a heterogeneous grassland at different periods (onset, mid and end) during summer. The performance of the two techniques was determined using the least relative root mean square error to the mean (nRMSEP) and the highest coefficients of determination (R^2_p) between the predicted and the measured LAI. Results show that iPLSR models could explain LAI variation with R^2_p values ranging from 0.81 to 0.93 and low nRMSEP from 9.39% to 24.71%. The highest accuracies for estimates of LAI using iPLSR were at mid- and end of summer (R^2_p = 0.93 and nRMSEP = 9.39%; R^2_p = 0.89 and nRMSEP = 10.50%, respectively). Pooling data sets from the three assessed periods yielded the highest prediction error (nRMSEP=24.71%). Results show that iPLSR performed better than PLSR, which yielded R^2_p and RMSEP values ranging from 0.36 to 0.65 and from 28.44% to 33.47%, respectively. Overall, this study demonstrates the value of iPLSR in predicting LAI and therefore provides a basis for more accurate mapping and monitoring of canopy characteristics of tropical grasslands.

Significance:

- The relationship between LAI and canopy reflectance can be used in iPLSR modelling to provide more accurate mapping and monitoring of canopy characteristics for land management and conservation.

Introduction

Measurement of spatio-temporal distribution of quantitative variables like leaf area index (LAI) and biomass are valuable for assessing the health and productivity of tropical grasslands.[1] Several studies (e.g. Prins and Beekman[2]) have associated vegetation characteristics such as LAI and biomass with animal grazing patterns. Therefore, quantitative assessment of such characteristics offers great potential for determining grassland conditions, which is useful for generating optimal management guidelines for grazing and rangeland conservation and restoration.

LAI has been recognised as a key biophysical parameter for determining vegetation characteristics.[3] LAI determines vegetation biophysical processes such as photosynthesis, canopy water interception, transpiration, radiation extinction, carbon loads and nutrient sequestration.[4,5] Consequently, LAI is commonly used as a key input for modelling vegetation foliage cover, growth and productivity and effects of disturbances such as drought and climate change on vegetation communities.[6]

Previous studies in which LAI was estimated on tropical grasslands have emphasised their spatial variation.[7] However, LAI is a biophysical parameter that is spatially and temporally dynamic across a landscape. According to Shen et al.[8], the performance of biophysical process models is highly sensitive to the temporal and spatial variation of LAI. For example, Xu and Baldocchi[9] note that well-timed data collection on changes in LAI could be used to explain more than 84% of the variance in gross primary production – an important input in the carbon cycle of an ecosystem. Therefore, analysis of temporal and spatial changes in LAI at the canopy level provides a valuable opportunity for modelling biophysical processes.

Traditionally, direct (e.g. destructive sampling) and indirect (e.g. use of a ceptometer canopy analyser and hemispherical canopy photography) methods are used to determine LAI in grasslands.[8,10,11] Typically, the direct methods consist of manually determining LAI using planimetric or volumetric techniques. Although these approaches are simple and reliable, they involve destructive sampling, are labour intensive, costly and time consuming.[1,12] These factors limit the application of direct methods for estimating LAI, particularly in large spatial extents that require frequent monitoring.[6] Indirect methods, like the use of a spectrometer, quantify LAI by measuring spectral reflectance which is then used as a proxy for LAI. Generally, such indirect methods are quick and can be automatically processed, thus allowing their application in a larger sampling area.[10]

Remotely sensed spectral data present an opportunity to indirectly retrieve LAI in heterogeneous grasslands.[1] Techniques that rely on remotely sensed spectral data are non-destructive, relatively quick and cost-effective, and therefore valuable for large spatial and multi-temporal monitoring.[8,13,14] The literature shows that canopy hyperspectral data, acquired using handheld spectrometers, have been widely adopted to derive LAI in heterogeneous grasslands.[15,16] According to Hansen and Schjoerring[16], such data provide hundreds or even thousands of spectral bands with information sensitive to specific vegetation variables valuable for modelling. Although Lee et al.[17] demonstrated that models generated from hyperspectral data predicted LAI better than those from broadband spectral data, the large amount of spectral information that characterises hyperspectral data makes derivation of LAI from heterogeneous grasslands data challenging.[7] Additionally, hyperspectral data sets suffer from multicollinearity that often occurs when many adjacent spectral bands present a high degree of redundancy and correlation.[18] Tropical grasslands

LAI retrieval using canopy reflectance is further complicated by varying species composition, phenology and proportions and complex canopy architecture.

A number of studies (e.g. Nguyen and Lee[19]) that have adopted canopy reflectance hyperspectral data to derive LAI have demonstrated the superiority of partial least square regression (PLSR) over traditional regression techniques. The technique was introduced to solve multicollinearity and overfitting problems by reducing variables to fewer components.[18] The PLSR technique is a full-spectrum method that simultaneously uses all available wavebands to create models. Compared to other algorithms, PLSR is less restrictive because it can be run on data for which the sample size is smaller than predictor variables.[20] The technique is particularly useful for removing uninformative bands and retaining those useful for predicting response variables. Consequently, it has become valuable for improving, inter alia, model predictions by reducing data collection costs, interpretation complexity and data dimensionality.[5,21] Moreover, PLSR combines the characteristics of popular statistical techniques such as stepwise mutiple regression and principal component regession. In several studies, PLSR turned out to be more robust than the regression techniques with which it was compared.[7,22,23] Furthermore, similar performance was found between radiative transfer and PLSR models in estimating LAI.[24]

Although the use of PLSR, a full-spectrum technique, has gained popularity in hyperspectral data modelling,[18,19,25] studies in fields like chemometrics have suggested that interval partial least squares (iPLSR), a variant of PLSR, can reduce hyperspectral data into band portions valuable for more accurate prediction.[26,27] Developed by Norgaard et al.[26], iPLSR is a graphically oriented technique for local regression modelling of spectral data. Unlike PLSR, it visually provides a general overview of relevant information in different spectral regions, thereby screening out important portions of the electromagnetic spectrum and discarding interference from irrelevant portions. Norgaard et al.[26], for instance, used spectra for beer samples to retrieve original extract concentration by comparing iPLSR, PLSR and other algorithms. They found that iPLSR improved determination coefficient and root mean square error of prediction of full-spectrum PLSR from 0.993 and 0.40% to 0.998 and 0.17%, respectively. Although this approach offers great promise in improving landscape modelling accuracy, no studies have used iPLSR on ground-based hyperspectral data collected from heterogeneous landscapes such as tropical grasslands.

To determine the value of specific spectral bands or regions to our models, we applied iPLSR to the entire electromagnetic spectrum. However, several studies have identified different spectral regions to relate to LAI variations. For example, Darvishzadeh et al.[7] and Zhao et al.[28] found that LAI-related bands were between near infrared (NIR) and short-wave infrared (SWIR) spectral regions. The same studies also noted that bands in the visible region (e.g. 440 nm) were valuable in LAI modelling. The relationship between LAI and red-edge bands has been established in several studies.[18,29,30] Generally, the value of a spectral band or region in estimating LAI depends on the vegetation status. For instance, at the senescence, the amount of chlorophyll drops, thus increasing the radiation of NIR and SWIR spectral bands and their contribution in modelling biochemical or biophysical parameters.[28] Consequently, we sought to pursue three objectives: (1) to identify useful bands for modelling LAI using iPLSR, (2) to compare heterogeneous tropical grasslands LAI estimates using iPLSR and PLSR models based on hyperspectral data and (3) to evaluate the robustness of the two models in estimating multi-temporal tropical grassland LAI (i.e. onset of, mid- and end summer) and pooled reflectance data during summer.

Materials and Methods

The study area

The study area was located in the Ukulinga Research Farm at the University of KwaZulu-Natal in Pietermaritzburg (30°24'S, 29°24'E) (Figure 1). The area is characterised by warm to hot summers and mild winters which often are accompanied by occasional frost. Mean monthly temperature ranges from 13.2 °C to 21.4 °C, with an annual mean of 17 °C.[31,32] The farm receives over 106 days of rain with an annual precipitation of about 680 mm. Soils originate from shallow marine shales of Lower Permian Ecca Group classified as Westleigh forms. The area is under the Southern Tall Grassveld and is predominately herbaceous as a result of frequent mowing and long-term burnings.[32] *Themeda triandra* Forssk, *Heteropogon contortus* (L.) P. Beauv. ex Roem. Schult. and *Tristachya leucothrix* Trin. ex Nees dominate the area.[33]

Field sampling

Data were collected during the southern hemisphere summer (October 2014 to March 2015). Stratified random sampling with clustering was adopted to select sampling sites. The grassland area was first digitised from an aerial photograph (Figure 1) and stratified into north, south, east and west aspects. To select the plots, 10 *x-y* coordinates were randomly generated from the stratum using the Hawth tool. In total, 40 plots (30 m x 30 m) were selected and located in the field using a GPS (Trimple GEO XT, with an estimated 100-mm accuracy). Two or three subplots of 1 m x 1 m were randomly chosen within each plot to generate a final sample size of 100 plots. Spectral and LAI data were then collected within the subplots at the onset of, mid- and end of summer.

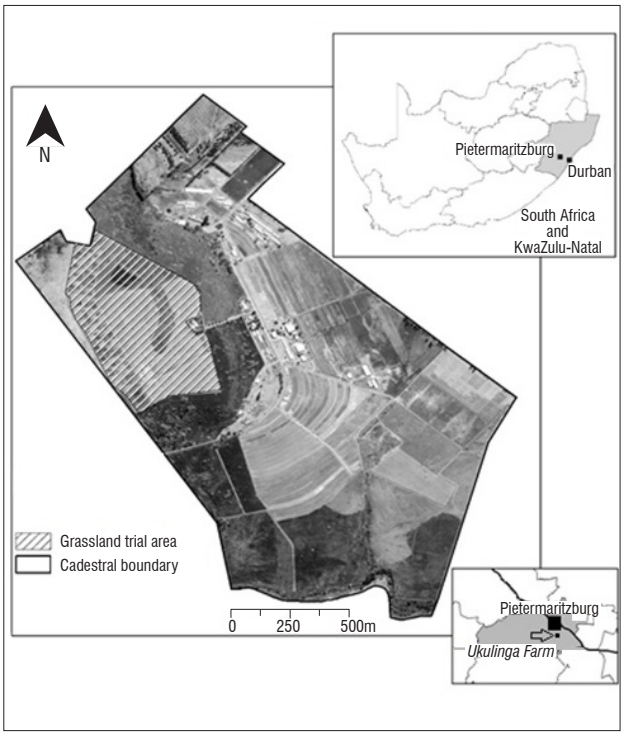


Figure 1: The Ukulinga Research Farm near the city of Pietermaritzburg in the province of KwaZulu-Natal, South Africa.

Data collection

At each sampling point, LAI was acquired with a LAI-2200C Plant Canopy Analyzer using the procedure described by Darvishzadeh et al.[7] Canopy reflectance was acquired using an analytical spectral device (ASD FieldSpec® 3 spectrometer, Boulder, CO, USA). The spectral resolution of the ASD FieldSpec® 3 spectrometer ranges from 350 nm to 2500 nm with 1.4-nm and 2-nm sampling intervals for the ultraviolet to visible and NIR region (350–1000 nm) and the SWIR region (1000–2500 nm) respectively. To normalise the spectra collected, the radiance of a white standard panel coated with barium sulfate and of known reflectivity was first recorded. Canopy reflectance measurements were made under clear sky between 10:00 and 14:00 local time to minimise atmospheric effects. To account for any changes in the atmospheric condition and the sun irradiance, reflectance measurements were recorded with frequent normalisation using the standard panel.[34] In total, 15 replicates of canopy reflectance within each subplot were collected and averaged, allowing for elimination of measurement noise arising from soil background.[7]

Data analysis

Pre-processing of hyperspectral data

To separate overlapping bands, thereby amplifying fine differences in the electromagnetic spectrum, the first-order derivative at three nanometres was applied on the resulting mean spectral data.[35,36] First-order derivative is also known to be useful in minimising atmospheric and background noise.[14,20] A number of researchers[7,37,38] have applied first-order derivative on hyperspectral data for LAI estimation. The spectral regions of 350–399 nm, 1355–1420 nm, 1810–1940 nm and 2470–2500 nm (Figure 2) are known to be noisy and were discarded from the spectra.[5,39]

Analysis of variance and Brown–Forsythe tests

The combined test of skewness and kurtosis was first employed to evaluate the distribution of the collected LAI data. The test of normality is a prerequisite to assessing data variability. A perfect normal distribution has skewness and kurtosis values equal to zero.[40] To assess LAI variations between periods within summer, one-way analysis of variance (ANOVA) and Brown–Forsythe tests ($\alpha = 0.05$) were implemented. The use of the Brown–Forsythe test, in addition to ANOVA, was justified by the smaller sample size at the end of summer ($n = 73$) because of spectrometer failure. According to Maxwell and Delaney[41] and Sheskin[42], the Brown–Forsythe test is preferred over ANOVA when sample sizes are heterogeneous and is less affected by data that are not normally distributed.

Partial least squares regression

Partial least squares regression was originally an econometric technique created by Herman Wold in the 1960s to construct predictive models from highly collinear explanatory variables.[25] The principle of PLSR is to firstly decompose explanatory variables (X) into a few non-correlated latent variables or components using information contained in the response variable (Y); then to regress the new components against the response variable.[23,43] According to Tan and Li[44], Wang et al.[45] and Yeniay and Goktas[25], the model that underlies PLSR consists of three phases. In the first phase, explanatory (X) and response (Y) variables are decomposed based on the expression:

$$X = TPT + E$$
$$Y = UQT + F, \qquad \text{Equation 1}$$

where T and U are respective matrices of scores of X and Y; P and Q stand for the matrices of loadings; and E and F for errors of X and Y matrices. In the second phase, the Y-scores (U) are predicted using the X-scores (T) based on the expression:

$$U = bT + e \qquad \text{Equation 2}$$

where b represents the regression coefficient and e the error matric of the relationship between Y-scores and X-scores. In the final phase, the predicted Y-scores are used to build predictive models of response variable using the expression:

$$Y = bTQ + G \qquad \text{Equation 3}$$

where G is the error matrix related to estimating Y.

In the present study, we used the PLS Toolbox (Eigenvector Research Inc.) with MATLAB (version R2013b) to build PLSR models. Before running PLSR, pre-processed hyperspectral data along with LAI data were autoscaled.[11] This procedure scales mean-centres of each waveband to unit standard deviation.[46] The PLSR was then run on data using a leave-one-out cross-validation method. The least root mean square error (RMSE) and the highest coefficients of determination (R^2) between the predicted and the measured Y variable were the two criteria used to select the best model with optimal number of components. The best model was suggested by the software.

Interval partial least squares regression

Interval partial least squares regression (iPLSR) is a variant of PLS that locally develops PLS models on equidistant portions of the full spectrum.[26,27] To predict a Y variable from spectra using iPLSR, the spectrum is split into a number of intervals of equal distance. A PLSR model is then built on each spectral interval. Thereafter, all the models built on the wavebands of different intervals are compared to the full-spectrum model based on calibration parameters such as root mean square error of cross-validation (RMSECV). Finally, the local model with the lowest RMSECV is selected.[21,47] The iPLSR can operate in two modes or variable selection directions: backward and forward mode. In forward mode, the algorithm starts without any variable selection and then develops the best PLSR model from the interval with the lowest RMSECV. This process can be repeated by including more intervals to enhance the model. In backward mode, the algorithm starts by selecting all variables and then discards the interval with the largest RMSECV.[48]

In this study, iPLSR in forward mode was used to select best spectral intervals. As predictive bands of LAI are known to spread across the entire electromagnetic spectrum as mentioned above, the interval size was set to a single variable. This approach is recommended when there is uniqueness of information in variables.[46] After several adjustments, the process was repeated 40 times. Therefore, the output local model had 40 intervals or bands. The iPLSR in forward mode was implemented using the PLS Toolbox.

Validation

A leave-one-out cross-validation method was implemented to calibrate models using 70% of the data and to find the optimal number of components. Then, the performance of trained models was validated using 30% of the data (independent data set). To assess model performance for prediction at the three sampling periods, relative root mean square regression to the mean (nRMSEP) and coefficient of determination (R^2_P) were used.

Data splitting into training and independent test data sets was performed using an onion algorithm.[43] An onion algorithm was chosen in this study to avoid arbitrary data splitting which may cause biased results.[7] The principle of onion algorithm is to keep outside covariant data plus those that are randomly inner spaced.[49]

Results

Variation in LAI and spectra data

The values of skewness (between 0.40 and -0.45) and kurtosis (between 0.86 and -0.11) indicate that the LAI of grass species canopy in the sampling plots had a normal distribution. Therefore the LAI data were suitable for the ANOVA and Brown–Forsythe tests. LAI variation in grass species canopy was significant among the three multi-temporal periods ($p < 0.01$). Samples in mid-summer had the highest mean (3.63 m^2/m^2) and variability (standard deviation = 1.10 m^2/m^2). Samples at the end of summer had the second highest mean (2.01 m^2/m^2) and lowest variability (standard deviation = 0.705 m^2/m^2). Samples at the beginning of summer had the least mean value of LAI (1.667 m^2/m^2) in grass species canopies, with the second least variability (0.821 m^2/m^2) in LAI.

To assess the change in reflectance at the different sampling periods, the mean spectra of all the sampling plots were averaged and upper and lower 95% confidence limits were derived. Results show that there was a change in averaged reflectance during the sampling periods (Figure 2). Visually, averaged reflectance was noticeably different across the electromagnetic spectrum. Canopy reflectance at the end, beginning and mid-summer presented the highest mean reflectance in the visible, NIR and SWIR regions, respectively. Figure 2 shows that first-derivative spectra differed in some spectral portions at the different sampling periods. The highest values of first-order derivative of reflectance are located in the NIR and SWIR region of the electromagnetic spectrum.

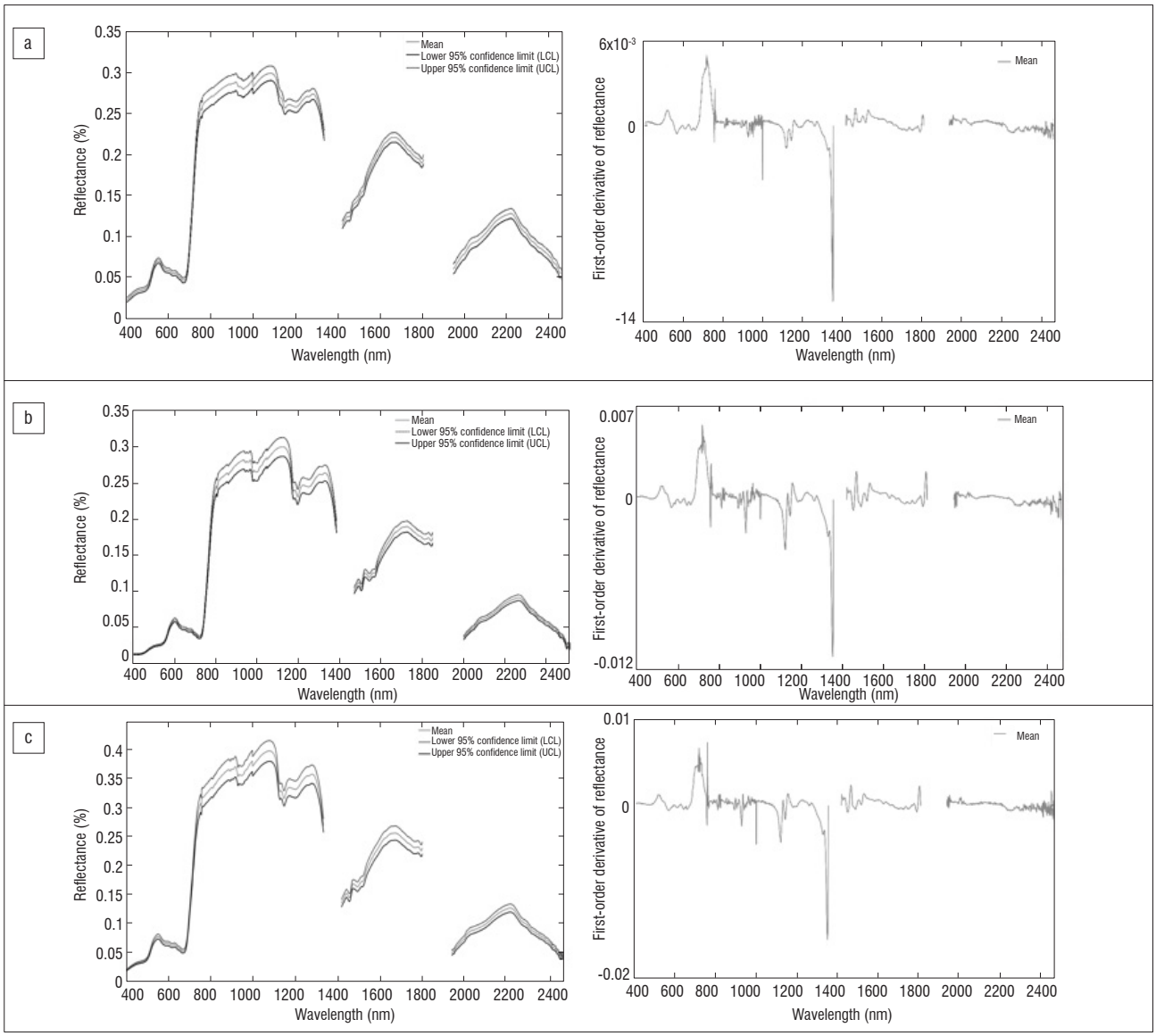


Figure 2: Mean and respective first-order derivative of canopy spectra of all grass subplots at the (a) beginning of, (b) mid- and (c) end of summer.

Table 1: R^2_{cv}, root mean square error (RMSE) and the number of components of training partial least square regression (PLSR) and interval PLSR (iPLSR) models prediction for the three sampling periods in summer and the pooled data

Regression algorithm	Number of components	R^2_{cv}	RMSE
Beginning of summer			
PLSR (full-spectrum)	6	0.31	0.74
iPLSR (40 intervals)	6	0.89	0.29
Middle of summer			
PLSR (full-spectrum)	4	0.54	0.77
iPLSR (40 intervals)	5	0.90	0.32
End of summer			
PLSR (full-spectrum)	5	0.39	0.55
iPLSR (40 intervals)	6	0.90	0.24
Pooled data			
PLSR (full-spectrum)	5	0.67	0.75
iPLSR (40 intervals)	6	0.81	0.53

PLSR and iPLSR models

Table 1 presents results of the model performance of PLSR and iPLSR for the training data set at each of the sampling periods within summer. Based on RMSECV and R^2, results show that the iPLSR models perform better than the PLSR models. At each period, iPLSR models were able to explain more than 85% of LAI variability (88.8% at the beginning, 90.3% of mid- and 89.6% at the end of summer) with RMSECV values that vary from 0.24 m^2/m^2 to 0.32 m^2/m^2. Although iPLSR had a slightly higher RMSECV value (0.53 m^2/m^2) it had a better estimation of LAI variability across the entire summer ($R^2_{cv} = 0.81$). PLSR models on the other hand yielded high RMSECV values (0.55–0.77 m^2/m^2) and poorly explained the LAI variation (31.3–67.1%).

The contribution of each waveband in the selected PLSR factors is displayed in Figure 3. The most valuable bands for estimating LAI were distributed across the electromagnetic spectrum. However, the highest peaks for all the periods within summer, including all the periods combined, were mostly located in the NIR and SWIR regions.

Using iPLSR models with 40 intervals, Table 2 and Figure 4 present the selected bands and their location within the four regions of the electromagnetic spectrum, respectively, while Figure 5 provides a percentage of predictive bands in relation to the regions within the electromagnetic spectrum.

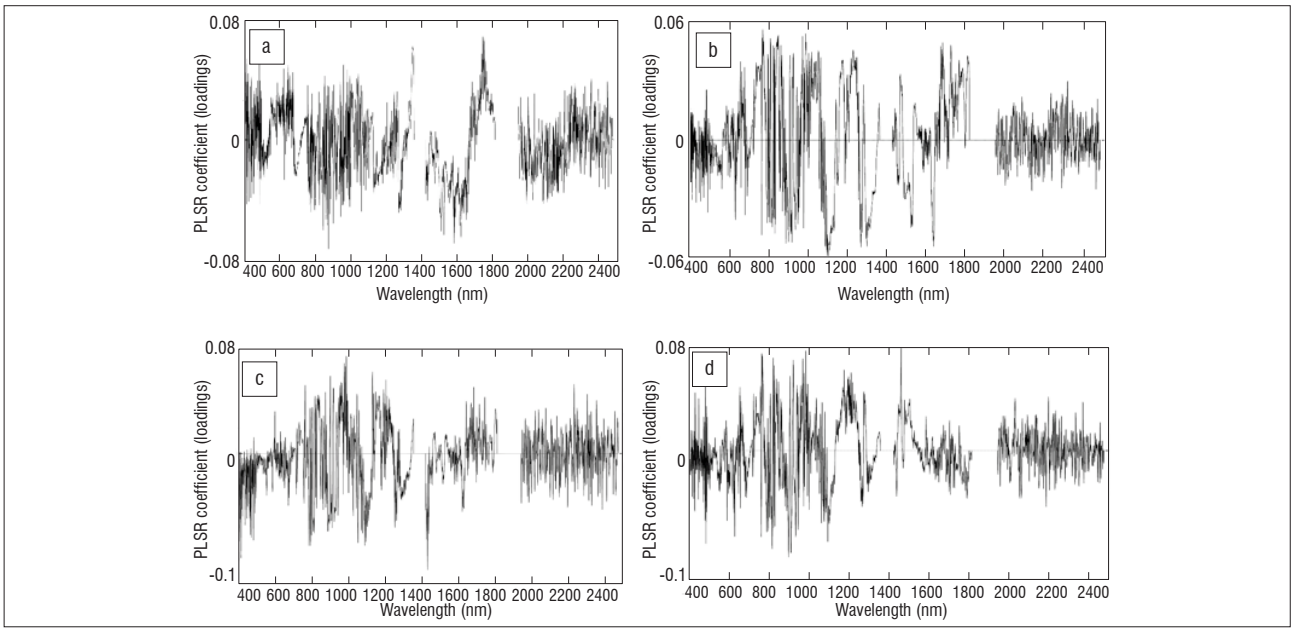
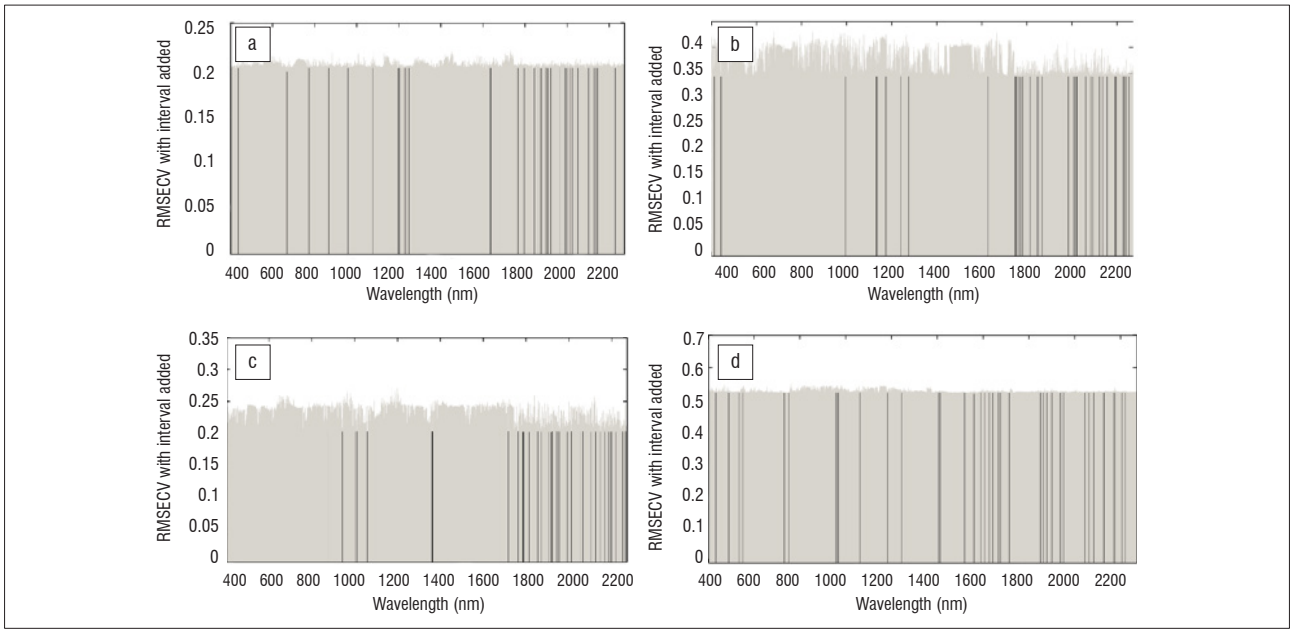


Figure 3: Partial least square regression (PLSR) loadings for (a) beginning of, (b) mid- and (c) end of summer and (d) pooled data.

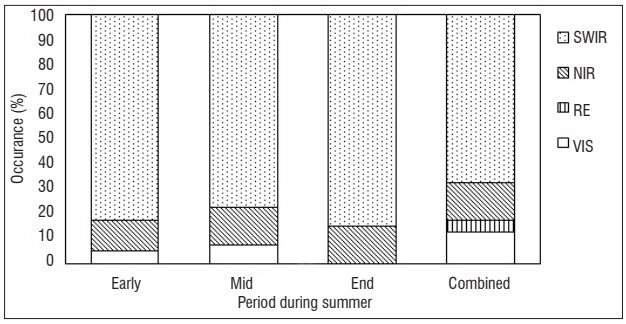
Table 2: Selected bands (nm) using interval partial least square regression models with 40 intervals for the three sampling periods in summer and the pooled data

	Visible	Red edge	Near infrared	Short-wave infrared
Beginning of summer	461, 764	–	793, 1020, 1061, 1201, 1267	1633, 1640, 1656, 1681, 1708, 1741, 1956, 1997, 2003, 2021, 2071, 2086, 2097, 2117, 2127, 2140, 2165, 2167, 2201, 2219, 2220, 2221, 2286, 2291, 2321, 2344, 2347, 2369, 2388, 2398, 2429, 2436, 2439
Mid-summer	413, 442, 443	–	995, 1132, 1134, 1174, 1240 , 1275	1693, 1944, 1947, 1951, 1959,1969, 1978, 2011, 2042, 2048, 2065, 2181, 2206, 2207, 2216, 2218, 2219, 2258, 2281, 2290, 2319, 2333, 2353, 2388, 2390, 2394, 2424, 2427 ,2434, 2437, 2450
End of summer	–	–	874, 943, 1003, 1010, 1058, 1059	1427, 1430, 1782, 1783, 1960, 1961, 1981, 1985, 1986, 2012, 2018, 2052, 2067, 2102, 2114, 2119, 2141, 2152, 2190, 2208, 2250, 2262, 2301, 2321, 2344, 2364, 2383, 2394, 2396, 2417, 2448, 2455, 2462, 2469
Pooled data	433, 489, 490, 535, 551	732, 752	957, 961, 968, 1062, 1183, 1244	1471, 1478, 1585, 1626, 1656, 1672, 1693, 1708, 1733, 1742, 1780, 2047, 2060, 2075, 2097, 2133, 2136, 2148, 2241, 2259, 2280, 2323, 2325, 2367, 2372, 2403, 2417



RMSECV, root mean square error of cross-validation

Figure 4: Optimal bands (in dark bars) selected by interval partial least square regression in developing leaf area index models at the (a) beginning of, (b) mid- and (c) end of summer and (d) pooled data.



SWIR, short-wave infrared; NIR, near infrared; RE, red edge; VIS, visible

Figure 5: Summary of predictive bands of leaf area index in different spectral regions.

Model validation

Figure 6 shows the performance of PLSR and iPLSR (40 intervals) models on the independent test data set. PLSR models of all the periods within summer (including all the periods combined) increased the coefficient of determination for prediction (R^2_p) and slightly decreased the relative root mean square error for prediction (nRMSEP). The values of R^2_p and nRMSEP, respectively, varied from 0.36 to 0.65 and from 28.44% (0.69 m^2/m^2) to 33.47% (0.56 m^2/m^2). However, iPLSR models performed better than the full-spectrum PLRS models for all the sampling periods in summer. The predictive power of iPLSR models did not change much on the validation data set. More than 80% of new data of LAI could be explained by the iPLSR models at all periods within summer (including all the periods combined).

Discussion

We sought to determine the performance of two multivariate regression models (PLSR and iPLSR) in estimating canopy level LAI on tropical grassland during summer. Comparisons were determined using the coefficient of determination (R^2) and the RMSE. Specifically, we examined the possibility of developing a model that can estimate LAI at different periods within summer (beginning, mid- and end) and across the entire summer period. Use of iPLSR to select the optimal bands for predicting LAI was also investigated.

Results showed that the PLSR algorithm run on first-derivative spectra to assess LAI variation at different periods did not perform well. The values of R^2_p and nRMSEP, respectively, ranged from 0.36 to 0.65 and 34.53% to 28.44%. Although PLSR is known to reduce the dimensionality of data to a few uncorrelated (orthogonal) components, inclusion of all the wavebands was not useful in the predictive performance of PLSR models – results consistent with Liu[50], Chung and Keles[51] and Filzmoser et al.[52] However, when data dimensionality was reduced to useful bands using iPLSR, the performance of models (R^2 and RMSE) significantly improved. Overall, there were very close relationships between measured and predicted LAI values, with low values of RMSE and higher values of determination coefficients (R^2) (Figure 6). Consistent with the findings of Zou et al.[53], Norgaard et al.[26] and Navea et al.[27], our findings confirm the superiority of iPLSR over full-spectrum PLSR.

The best predictive performance was derived from canopy reflectance at mid- ($R^2_p = 0.93$ and nRMSEP = 9.39%) and end summer ($R^2_p = 0.89$ and nRMSEP = 10.50%). The models performed the worst at the beginning of summer ($R^2_p = 0.88$ and nRMSEP = 17.37%) and for all the sampling periods combined ($R^2_p = 0.81$ and nRMSEP = 24.71%). The lower early summer prediction in comparison to the two other sampling periods can be attributed to higher soil background noise. According to Darvishzadeh et al.[7], soil background often has a negative effect on the predictive power of hyperspectral data when LAI is low. The lower performance at the end of summer in comparison to mid-summer might also be caused by soil background reflectance emanating from litters.

Adoption of iPLSR was useful in identifying relevant wavebands for predicting LAI. In total, 40 intervals were identified for all the sampling periods. The success of iPLSR for band selection in this study may be attributed to successful separation of overlapping bands performed by the first-derivative technique on the spectra. The spectral regions (NIR and SWIR) of bands selected by iPLSR are consistent with the findings by Darvishzadeh et al.[7], Thenkabail et al.[38], Brown et al.[54] and Gong et al.[55] Within ±12 nm, the bands chosen (Figure 4) in this study showed a consistency with the known bands for estimating LAI. For example, bands near 793 nm, 1061 nm, 1062 nm, 1633 nm, 442 nm, 443 nm, 535 nm, 551 nm, 732 nm and 2190 nm were also identified by Wang et al.[37] for estimating rice LAI at different growth phases. Furthermore, Gong et al.[55] found that bands centred near 1201 nm, 1240 nm, 1062 nm, 1640 nm, 2097 nm and 2259 nm were useful for estimating forest LAI.

It is worth noting that the contribution of different spectral regions along with their wavebands to LAI estimation depends on a particular period within summer (Figure 4). This dependence might be explained by the fact that the positions of selected wavebands are sensitive to changes in LAI, as indicated by ANOVA and Brown–Forsythe test results. Thus, the positions vary when factors like biochemical (e.g. chlorophyll) and biophysical (e.g. canopy closure) parameters and background effects change with canopy growth phases.[37] For example, at the end of summer, as the canopy senesces and the amount of chlorophyll declines, NIR and SWIR become more important in predicting LAI.[28] Furthermore, in the combined period, the selected bands can be explained by the fact that they were insensitive to changes in LAI (see Table 2). Delegido et al.[56] found that vegetation indices combining bands at 674 nm and 712 nm could overcome the aforementioned saturation problem while Kim et al.[57] found similar results with the ratio of 550 nm and 700 nm, which were insensitive to changes in chlorophyll concentration.

In this study, iPLSR models have proved to outperform full-spectrum PLSR models. However, model performance has shown to depend on the period within summer, on vegetation and on site conditions. These limitations are expected because PLSR and its variants (e.g. iPLSR), which are linear regression techniques, empirically relate to LAI and spectral reflectance, which makes the models non-transferable when environmental conditions of grassland (or vegetation cover in general) change.[24] Further work should look at comparing iPLSR with other robust and flexible methods, such as physically based radiative transfer models, particularly for the combined period. Models for the combined period used physical laws to explicitly relate biophysical variables and spectral variation of canopy reflectance. Consequently, these models are known to be more reproducible than linear regression models such as PLSR.[58] Currently, rapid development is being undertaken on physically based radiative transfer models for application in the field of remote sensing.[59] Further studies should also compare iPLSR with non-linear machine learning (e.g. random forest, support vector machine) techniques as they are able to cope with non-linear relationships between biophysical variables and canopy reflectance in dense grasslands.[60]

Conclusions

The following conclusions can be drawn:

- iPLSR can be used to simplify the relationship between LAI and canopy reflectance transformed using first-derivative technique better than PLSR can. The best iPLSR relationship is at the beginning and end of summer.
- By including all the variables, full-spectrum PLSR models yield a higher prediction error.
- iPLSR used as a single variable selection algorithm for LAI estimation can generate stable and reliable models with 40 bands.
- The period within summer, which is associated with vegetation growth, determines the selection and accuracy of LAI predictive bands.

Results show that appropriate band selection on in-situ hyperspectral data using iPLSR can overcome the challenge faced by remotely sensed data to accurately estimate LAI in a heterogeneous grassland. The findings pave the way to more accurate mapping and monitoring of canopy characteristics in a tropical grassland from airborne and space-borne hyperspectral data. However, the development of a iPLSR model for all the periods combined within summer needs further investigation, as its prediction error was higher than those for the periods separately.

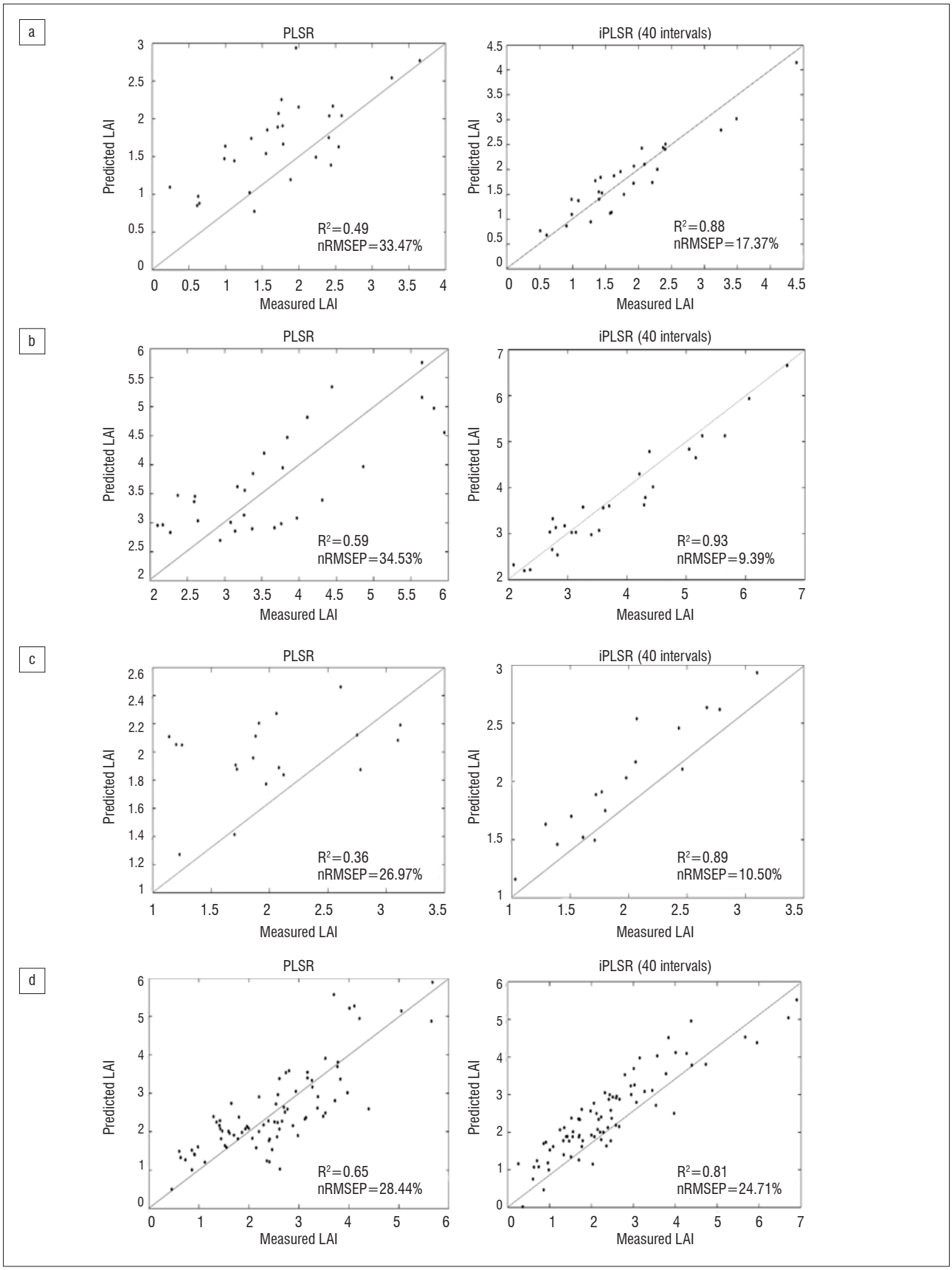


Figure 6: One-to-one relationship (m^2/m^2) between measured and predicted leaf area index (LAI) for validating partial least square regression (PLSR) and interval partial least square regression (iPLSR) models on an independent test data set in (a) early summer, (b) mid-summer, (c) end of summer and for the (d) pooled data.

Acknowledgements

We thank Brice Gijsbertsen and Victor Bangamwabo for technical support. We thank the Eigenvector Research Inc. for the free software PLS Toolbox. We also are grateful to Ognelet Marie Claude for his financial support throughout the life cycle of this research. Lastly, we acknowledge fellow students Dube Thimothy and Mfundiso Cele for their field assistance and the anonymous reviewers for improving the quality of this paper.

Authors' contributions

Z.K. was responsible for the data analysis and write-up; J.O. and O.M. edited and revised the manuscript.

References

1. He Y, Guo X, Wilmshurst JF. Comparison of different methods for measuring leaf area index in a mixed grassland. Can J Plant Sci. 2007;87(4):803–813. https://doi.org/10.4141/CJPS07024
2. Prins HHT, Beekman JH. A balanced diet as a goal for grazing – the food of the manyara buffalo. Afr J Ecol. 1989;27(3):241–259. https://doi.org/10.1111/j.1365-2028.1989.tb01017.x
3. Broge NH, Mortensen JV. Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data. Remote Sens Environ. 2002;81(1):45–57. https://doi.org/10.1016/S0034-4257(01)00332-7
4. Chen JM, Cihlar J. Retrieving leaf area index of boreal conifer forests using Landsat TM images. Remote Sens Environ. 1996;55(2):153–162. https://doi.org/10.1016/0034-4257(95)00195-6
5. Abdel-Rahman EM, Mutanga O, Odindi J, Adam E, Odindo A, Ismail R. A comparison of partial least squares (PLS) and sparse PLS regressions for predicting yield of Swiss chard grown under different irrigation water sources using hyperspectral data. Comput Electron Agric. 2014;106:11–19. https://doi.org/10.1016/j.compag.2014.05.001
6. Breda NJJ. Ground-based measurements of leaf area index: A review of methods, instruments and current controversies. J Exp Bot. 2003;54(392):2403–2417. https://doi.org/10.1093/jxb/erg263
7. Darvishzadeh R, Skidmore A, Schlerf M, Atzberger C, Corsi F, Cho M. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. ISPRS J Photogramm Remote Sens. 2008;63(4):409–426. https://doi.org/10.1016/j.isprsjprs.2008.01.001
8. Shen L, Li Z, Guo X. Remote sensing of leaf area index (LAI) and a spatiotemporally parameterized model for mixed grasslands. Int J Appl. 2014;4(1):46–61.
9. Xu LK, Baldocchi DD. Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California. Agric For Meteorol. 2004;123(1–2):79–96. https://doi.org/10.1016/j.agrformet.2003.10.004
10. Jonckheere I, Fleck S, Nackaerts K, Muys B, Coppin P, Weiss M, et al. Review of methods for in situ leaf area index determination – Part I. Theories, sensors and hemispherical photography. Agric For Meteorol. 2004;121(1–2):19–35. https://doi.org/10.1016/j.agrformet.2003.08.027
11. Zhang R, Ba J, Ma Y, Wang S, Zhang J, Li W, editors. A comparative study on wheat leaf area index by different measurement methods. Proceedings of the First International Conference on Agro-Geoinformatics; 2012 August 2–4: Shangai, China. IEEE; 2012. https://doi.org/10.1109/Agro-Geoinformatics.2012.6311671
12. Chason JW, Baldocchi DD, Huston MA. A comparison of direct and indirect methods for estimating forest canopy leaf-area. Agric For Meteorol. 1991;57(1–3):107–128. https://doi.org/10.1016/0168-1923(91)90081-Z
13. Bulcock HH, Jewitt GPW. Spatial mapping of leaf area index using hyperspectral remote sensing for hydrological applications with a particular focus on canopy interception. Hydrol Earth Syst Sci. 2010;14(2):383–392. https://doi.org/10.5194/hess-14-383-2010
14. Pullanagari RR, Yule IJ, Tuohy MP, Hedley MJ, Dynes RA, King WM. In-field hyperspectral proximal sensing for estimating quality parameters of mixed pasture. Precis Agric. 2012;13(3):351–369. https://doi.org/10.1007/s11119-011-9251-4
15. Atzberger C, Jarmer T, Schlerf M, Kötz B, Werner W, editors. Spectroradiometric determination of wheat bio-physical variables. Comparison of different empirical-statistical approaches. In: Remote Sensing in Transition; 2003 June 2–5; Ghent, Belgium. Rotterdam: Millpress; 2004. Available from: http://www.geo.uzh.ch/microsite/rsl-documents/research/publications/other-sci-communications/Atzberger_etal_Gent2003-2987622144/Atzberger_etal_Gent2003.pdf
16. Hansen P, Schjoerring J. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote Sens Environ. 2003;86(4):542–553. https://doi.org/10.1016/S0034-4257(03)00131-7
17. Lee KS, Cohen WB, Kennedy RE, Maiersperger TK, Gower ST. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. Remote Sens Environ. 2004;91(3–4):508–520. https://doi.org/10.1016/j.rse.2004.04.010
18. Li X, Zhang Y, Bao Y, Luo J, Jin X, Xu X, et al. Exploring the best hyperspectral features for LAI estimation using partial least squares regression. Remote Sens. 2014;6(7):6221–6241. https://doi.org/10.3390/rs6076221
19. Nguyen HT, Lee BW. Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression. Eur J Agron. 2006;24(4):349–356. https://doi.org/10.1016/j.eja.2006.01.001
20. Dorigo WA, Zurita-Milla R, De Wit AJW, Brazile J, Singh R, Schaepman ME. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. Int J Appl Earth Obs Geoinf. 2007;9(2):165–193. https://doi.org/10.1016/j.jag.2006.05.003
21. Andersen CM, Bro R. Variable selection in regression-a tutorial. J Chemom. 2010;24(11–12):728–737. https://doi.org/10.1002/cem.1360
22. Atzberger C, Guérif M, Baret F, Werner W. Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. Comput Electron Agric. 2010;73(2):165–173. https://doi.org/10.1016/j.compag.2010.05.006
23. Cho MA, Skidmore A, Corsi F, Van Wieren SE, Sobhan I. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. Int J Appl Earth Obs Geoinf. 2007;9(4):414–424. https://doi.org/10.1016/j.jag.2007.02.001
24. Darvishzadeh R, Atzberger C, Skidmore A, Schlerf M. Mapping grassland leaf area index with airborne hyperspectral imagery: A comparison study of statistical approaches and inversion of radiative transfer models. ISPRS Int J Remote Sens. 2011;66(6):894–906. https://doi.org/10.1016/j.isprsjprs.2011.09.013
25. Yeniay O, Goktas A. A comparison of partial least squares regression with other prediction methods. Hacet J Math Stat. 2002;31(99):99–101.
26. Norgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. Appl Spectrosc. 2000;54(3):413–419. https://doi.org/10.1366/0003702001949500
27. Navea S, Tauler R, De Juan A. Application of the local regression method interval partial least-squares to the elucidation of protein secondary structure. Anal Biochem. 2005;336(2):231–242. https://doi.org/10.1016/j.ab.2004.10.016
28. Zhao D, Huang L, Li J, Qi J. A comparative analysis of broadband and narrowband derived vegetation indices in predicting LAI and CCD of a cotton canopy. ISPRS J Photogramm Remote Sens. 2007;62(1):25–33. https://doi.org/10.1016/j.isprsjprs.2007.01.003
29. Mutanga O, Skidmore AK. Narrow band vegetation indices overcome the saturation problem in biomass estimation. Int J Remote Sens. 2004;25(19):3999–4014. https://doi.org/10.1080/01431160310001654923
30. Pu R, Gong P, Biging GS, Larrieu MR. Extraction of red edge optical parameters from hyperion data for estimation of forest leaf area index. IEEE Transactions on Geoscience and Remote Sensing. 2003;41(4):916–921. https://doi.org/10.1109/TGRS.2003.813555
31. Everson CS, Mengistu MG, Gush MB. A field assessment of the agronomic performance and water use of *Jatropha curcas* in South Africa. Biomass Bioenerg. 2013;59:59–69. https://doi.org/10.1016/j.biombioe.2012.03.013
32. Mills AJ, Fey MV. Frequent fires intensify soil crusting: Physicochemical feedback in the pedoderm of long-term burn experiments in South Africa. Geoderma. 2004;121(1–2):45–64. https://doi.org/10.1016/j.geoderma.2003.10.004

33. Ghebrehiwot HM, Kulkarni MG, Szalai G, Soos V, Balazs E, Van Staden J. Karrikinolide residues in grassland soils following fire: Implications on germination activity. S Afr J Bot. 2013;88:419–424. https://doi.org/10.1016/j.sajb.2013.09.008

34. Rajah P, Odindi J, Abdel-Rahman EM, Mutanga O, Modi A. Varietal discrimination of common dry bean (*Phaseolus vulgaris* L.) grown under different watering regimes using multi-temporal hyperspectral data. J Appl Remote Sensing. 2015;9(1):096050–096050.

35. Archontaki HA, Atamian K, Panderi IE, Gikas EE. Kinetic study on the acidic hydrolysis of lorazepam by a zero-crossing first-order derivative UV-spectrophotometric technique. Talanta. 1999;48(3):685–693. https://doi.org/10.1016/S0039-9140(98)00288-4

36. Holden H, LeDrew E. Spectral discrimination of healthy and non-healthy corals based on cluster analysis, principal components analysis, and derivative spectroscopy. Remote Sens Environ. 1998;65(2):217–224. https://doi.org/10.1016/S0034-4257(98)00029-7

37. Wang F-m, Huang J-f, Zhou Q-f, Wang X-z. Optimal waveband identification for estimation of leaf area index of paddy rice. J Zhejiang Univ Sci B. 2008;9(12):953–963. https://doi.org/10.1631/jzus.B0820211

38. Thenkabail PS, Enclona EA, Ashton MS, Van der Meer B. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. Remote Sens Environ. 2004;91(3–4):354–376. https://doi.org/10.1016/j.rse.2004.03.013

39. Adjorlolo C, Mutanga O, Cho MA, Ismail R. Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification. Int. J Appl Earth Obs. Geoinf. 2013;21:535–544. https://doi.org/10.1016/j.jag.2012.07.011

40. Peat J, Barton B. Medical statistics: A guide to data analysis and critical appraisal. Malden, MA: Blackwell Publishing; 2005. https://doi.org/10.1002/9780470755945

41. Maxwell SE, Delaney HD. Designing experiments and analyzing data: A model comparison perspective. 2nd ed. New York: Psychology Press; 2004.

42. Sheskin DJ. Handbook of parametric and nonparametric statistical procedures. Boca Raton, FL: Chapman and Hall/CRC; 2011.

43. Tobias RD, editor. An introduction to partial least squares regression. Paper presented at: Twentieth Annual SAS Users Group International conference; 1995 April 2–5; Orlando, Florida, USA.

44. Tan C, Li M. Mutual information-induced interval selection combined with kernel partial least squares for near-infrared spectral calibration. Acta Mol Biomol Spectrosc. 2008;71(4):1266–1273. https://doi.org/10.1016/j.saa.2008.03.033

45. Wang F-m, Huang J-f, Lou Z-h. A comparison of three methods for estimating leaf area index of paddy rice from optimal hyperspectral bands. Precis Agric. 2011;12(3):439–447. https://doi.org/10.1007/s11119-010-9185-2

46. Wise BM, Gallagher NB, Bro R, Shaver JM, Windig W, Koch RS. PLS_Toolbox version 4.0 for use with MATLAB™. Manson, WA: Eigenvector; 2006

47. Bezerra de Lira LF, De Albuquerque MS, Andrade Pacheco JG, Fonseca TM, De Siqueira Cavalcanti EH, Stragevitch L, et al. Infrared spectroscopy and multivariate calibration to monitor stability quality parameters of biodiesel. Microchem J. 2010;96(1):126–131. https://doi.org/10.1016/j.microc.2010.02.014

48. Mehmood T, Liland KH, Snipen L, Saebo S. A review of variable selection methods in partial least squares regression. Chemometr Intell Lab. 2012;118:62–69. https://doi.org/10.1016/j.chemolab.2012.07.010

49. Sousa AG, Ahl LI, Pedersen HL, Fangel JU, Sorensen SO, Willats WGT. A multivariate approach for high throughput pectin profiling by combining glycan microarrays with monoclonal antibodies. Carbohydr Res. 2015;409:41–47. https://doi.org/10.1016/j.carres.2015.03.015

50. Liu J. Developing a soft sensor based on sparse partial least squares with variable selection. J Process Contr. 2014;24(7):1046–1056. https://doi.org/10.1016/j.jprocont.2014.05.014

51. Chung D, Keles S. Sparse partial least squares classification for high dimensional data. Stat Appl Genet Mol. 2010;9(1), Art. #1492. https://doi.org/10.2202/1544-6115.1492

52. Filzmoser P, Gschwandtner M, Todorov V. Review of sparse methods in regression and classification with application to chemometrics. J Chemom. 2012;26(3–4):42–51. https://doi.org/10.1002/cem.1418

53. Zou X, Zhao J, Huang X, Li Y. Use of FT-NIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of 'Fuji' apple based on different PLS models. Chemometr Intell. 2007;87(1):43–51. https://doi.org/10.1016/j.chemolab.2006.09.003

54. Brown L, Chen JM, Leblanc SG, Cihlar J. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: An image and model analysis. Remote Sens Environ. 2000;71(1):16–25. https://doi.org/10.1016/S0034-4257(99)00035-8

55. Gong P, Pu RL, Biging GS, Larrieu MR. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. IEEE Trans Geosci Remote Sens. 2003;41(6):1355–1362. https://doi.org/10.1109/TGRS.2003.812910

56. Delegido J, Verrelst J, Meza CM, Rivera JP, Alonso L, Moreno J. A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. Eur J Agron. 2013;46:42–52. https://doi.org/10.1016/j.eja.2012.12.001

57. Kim MS, Daughtry CST, Chappelle EW, Mcmurtrey JE, Walthall CL, editors. The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (A par). In: CNES, Proceedings of 6th International Symposium on Physical Measurements and Signatures in Remote Sensing; 1994; Val D'Isere, France. Val D'Isere: The Symposium, 1994. p. 299–306.

58. Quan X, He B, Yebra M, Yin C, Liao Z, Zhang X, et al. A radiative transfer model-based method for the estimation of grassland aboveground biomass. Int J Appl Earth Obs Geoinf. 2017;54:159–168. https://doi.org/10.1016/j.jag.2016.10.002

59. Jacquemoud S, Verhoef W, Baret F, Bacour C, Zarco-Tejada PJ, Asner GP, et al. PROSPECT+ SAIL models: A review of use for vegetation characterization. Remote Sens Environ. 2009;113:S56–S66. https://doi.org/10.1016/j.rse.2008.01.026

60. Kiala Z, Odindi J, Mutanga O, Peerbhay K. Comparison of partial least squares and support vector regressions for predicting leaf area index on a tropical grassland using hyperspectral data. J Appl Remote Sens. 2016;10(3), Art. #036015, 14 pages. https://doi.org/10.1117/1.JRS.10.036015

The benefits of segmentation: Evidence from a South African bank and other studies

AUTHORS:
Douw G. Breed[1]
Tanja Verster[1]

AFFILIATION:
[1]Centre for Business Mathematics and Informatics, North-West University, Potchefstroom, South Africa

CORRESPONDENCE TO:
Tanja Verster

EMAIL:
Tanja.Verster@nwu.ac.za

DATES:
Received: 10 Nov. 2016
Revised: 09 May 2017
Accepted: 15 May 2017

KEYWORDS:
predictive models; case studies; logistic regression; linear modelling; semi-supervised segmentation

HOW TO CITE:
Breed DG, Verster T. The benefits of segmentation: Evidence from a South African bank and other studies. S Afr J Sci. 2017;113(9/10), Art. #2016-0345, 7 pages. http://dx.doi.org/10.17159/sajs.2017/20160345

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Department of Science and Technology (South Africa)

We applied different modelling techniques to six data sets from different disciplines in the industry, on which predictive models can be developed, to demonstrate the benefit of segmentation in linear predictive modelling. We compared the model performance achieved on the data sets to the performance of popular non-linear modelling techniques, by first segmenting the data (using unsupervised, semi-supervised, as well as supervised methods) and then fitting a linear modelling technique. A total of eight modelling techniques was compared. We show that there is no one single modelling technique that always outperforms on the data sets. Specifically considering the direct marketing data set from a local South African bank, it is observed that gradient boosting performed the best. Depending on the characteristics of the data set, one technique may outperform another. We also show that segmenting the data benefits the performance of the linear modelling technique in the predictive modelling context on all data sets considered. Specifically, of the three segmentation methods considered, the semi-supervised segmentation appears the most promising.

Significance:

- The use of non-linear modelling techniques may not necessarily increase model performance when data sets are first segmented.
- No single modelling technique always performed the best.
- Applications of predictive modelling are unlimited; some examples of areas of application include database marketing applications; financial risk management models; fraud detection methods; medical and environmental predictive models.

Introduction

Predictive modelling is the general concept of building a model that is capable of making predictions by predicting a target variable based on various explanatory variables. Specifically in this paper, the target variable will be binary, i.e. there are only two possible outcomes.

The number of modelling techniques available in predictive modelling is extensive.[1] These techniques can be split into linear and non-linear modelling techniques. Linear modelling techniques assume a linear relationship between the target variable and each explanatory variable. Linear modelling techniques are typically easier to understand and very transparent. For these reasons, linear modelling techniques are the most used techniques in industry. However, linear modelling techniques may in some cases perform worse in terms of model performance and may be less robust as a result of the linearity assumption made. In this paper, we show that, by first segmenting the data, linear modelling techniques can perform just as well (and sometimes better) than popular non-linear modelling techniques.

Non-linear modelling techniques, on the other hand, are typically more complex and do not assume a linear relationship between the target variable and each explanatory variable. Non-linear modelling techniques are not as transparent but usually more robust and sometimes perform better in terms of model performance.[2]

In the process of determining how well a predictive modelling technique performs, the lift of the model is considered, where lift is defined as the ability of a model to distinguish between the two outcomes of the target variable.[3] There are several ways to measure model lift and in this paper Gini coefficient was chosen.

Segmentation of the data that are used for predictive modelling is a well-established practice in the industry.[4-6] The ultimate goal of any segmentation (in the predictive modelling context) is to achieve more accurate, robust and transparent models.[6] Segmentation is defined as the practice of classifying (or partitioning) data observations into distinct groups or subsets with the aim of developing predictive models on each of the groups separately, in order to improve the overall predictive power.

Two main streams of statistical segmentation exist in the industry, namely unsupervised and supervised segmentation.[7,8] Unsupervised segmentation[7] focuses on the explanatory variables in the models, whereas supervised segmentation[8] focuses on the target variable. We also used a third stream, combining both aspects, called semi-supervised segmentation, as developed in a recent PhD thesis.[9]

The main objective of this paper was to compare the model performance when first segmenting the data before fitting a linear modelling technique to the model performance of popular non-linear modelling techniques that may not require segmentation. Of the three methods of segmentation that were compared, semi-supervised segmentation looks the most promising overall.

Modelling techniques

Linear modelling technique

The most common linear modelling technique is linear regression; however, when modelling a binary target variable using linear regression, two problems arise.[2] The first problem is that one of the assumptions underlying the linear regression model does not hold, namely normally distributed error terms. The second problem is that in linear regression, no bounds are on the target variable, whereas with a binary target variable, the target variable is restricted to two outcomes. To overcome these problems, logistic regression is used (by combining linear regression with a specific bounding function), which is sometimes referred to as the logit transformation[10], i.e. the log of the odds of the probability of the target variable.

Technical specifications

SAS software's Proc Logistic was used with default settings with the addition of using stepwise selection as a subset selection criterion. Note that this means that the final regression analysis does not use all of the explanatory variables, but selects a subset of variables that explains the target variable in the most efficient way.

Three methods of segmentation

We first segmented the data before fitting a logistic regression to the data. As mentioned, two main streams of statistical segmentation exist in the industry: unsupervised and supervised segmentation.[7,8] Unsupervised segmentation[7] focuses on the explanatory variables in the models to be developed and does not take the target variable into account; a popular example of unsupervised segmentation is clustering. Supervised segmentation focuses on the target variable; a popular example of supervised segmentation is the decision tree.

Both these streams make intuitive sense depending on the application and the requirements of the models developed[11] and many examples exist in which the use of either technique has improved model performance[12]. However, both these streams focus on a single aspect (i.e. either target separation or independent variable distribution) and combining both aspects might better deliver. This approach is explored in Breed et al.[13] and described in more detail in a recent PhD thesis[9] and was used as the third segmentation method. This specific technique uses k-means clustering to measure the independent variable distribution and uses information value to measure target separation. A supervised weight is defined to measure the balance between the two aspects.[13] This algorithm is thus called SSSKMIV to indicate semi-supervised segmentation as applied to k-means using information value. The implementation of this algorithm is quite complex and the detail can be found in Breed[9].

Technical specifications

All three segmentation methods were implemented in SAS. The detail of the technical specifications (e.g. the optimal number of segments, the weight parameters in SSSKMIV, the optimal value of k in the k-mean algorithm, heuristic example) can be found in Breed[9].

Non-linear modelling techniques

For the non-linear modelling techniques, we used: neural networks; support vector machines; memory-based reasoning; decision trees (used here as the final model, not as segmentation) and gradient boosting (which is a boosting variation of random forests).

Technical specifications

The results of the non-linear modelling techniques were obtained through SAS Enterprise Miner software using specific nodes. Nodes are tools in SAS Enterprise Miner that implement, for example, different modelling techniques.[14] By using the default settings in SAS Enterprise Miner, the benefits of most techniques were utilised (e.g. subset selection is automatically done and complexity is automatically optimised).

Neural networks (also known as multilayer perceptrons) are often regarded as mysterious and powerful predictive tools, but on closer inspection the most typical form of a neural network is just a regression model with a flexible addition. The power of this addition must not be underestimated and enables the neural network to model virtually any relationship between the explanatory variable and the target variable.[14] Neural networks have been researched since the early 1940s[15] and are very well known in the predictive modelling field today.

Technical specifications

The AutoNeural Node of SAS Enterprise Miner software was used with default settings.

Support vector machines were introduced in the 1990s and are considered to be relatively new (compared to other well-known modelling techniques).[16] Support vector machines have been researched quite extensively over the last number of years.[17-19]. They predict a binary target by maximising the margin between the two outcomes through hyperplanes; more detail can be found in Meyer and Wien[19].

Technical specifications

The Enterprise Miner SVM (support vector machine) node was used with default settings, with one exception: the estimation method was set to least squares support vector machine as opposed to decomposed quadratic programming, as this setting failed to find a conclusive result on one of the six data sets (the claim prediction data set).

Memory-based reasoning uses k-nearest-neighbour principles to classify observations in a data set. When a new observation is evaluated, the algorithm allows the k-nearest observations of the development set to 'vote' regarding that observation's classification (their votes are based on the values of their target variables). These votes then represent the probabilities of the new observation belonging to that specific target value.

Technical specifications

The memory-based reasoning node was used with the default settings provided by the SAS Enterprise Miner software.

Decision trees are simple classifiers that produce prediction rules that are easy to interpret and apply and are commonly referred to as CART (classification and regression trees). For this reason they are also quite popular in the industry.[4] Note that usually decision trees are used to segment data as an example of supervised segmentation, but here decision trees are used for predictive modelling.

Technical specifications

The decision tree node in SAS Enterprise Miner was used. Two changes were made to the default settings. The splitting criterion for nominal input variables was changed from chi-squared probability to Gini, as the Gini is the measure we used for model performance in this paper. In addition, the number of branches or subsets that a splitting rule can produce was increased to six, which allows results that are more granular.

Gradient boosting draws its concept from the greedy decision tree approach proposed by Friedman[20]. The algorithm creates a number of small decision trees on the development set, and these trees are combined to produce the model's output. The technique can be linked to the techniques used in random forests[21] in that a number of different trees are developed[7].

Technical specifications

The gradient boosting node was used with its default setting in Enterprise Miner.

Model performance

In order to compare the model performance, each data set was first divided into two equal sets to form a development and a validation set. The development set, sometimes referred to as the training data, is used to develop the predictive models, whilst the validation set, alternatively known as the holdout data, is used to test the lift in model performance as measured by the Gini coefficient (hereafter lift).[5] The Gini coefficient

therefore quantifies the ability of the model to differentiate between the two outcomes of the target variable.[3] Obviously many other performance measures could have been used. The Gini coefficient is one of the most popular measures to use in retail credit scoring[4-6] and has the added advantage that it is a single number[3].

The development set and validation set were randomly sampled with even sizes (i.e. 50% each). Although the norm is to usually use larger samples for the development set (70–80%, resulting in a 20–30% validation set), the validation Gini is used in this paper as the ultimate measure of success, and the larger validation set size of 50% was therefore preferred to ensure the Gini coefficients are not affected by low sample size.

In order to measure the combined Gini of the segmented models on the validation set, the predicted probabilities of all segments were combined, and the Gini was calculated on the overall, combined set.

In summary, the eight modelling techniques are shown in Table 1.

Data sets

The modelling techniques described above were compared on six different data sets. All explanatory variables were standardised (i.e. by subtracting the mean and dividing by the standard deviation). Standardising data is a data pre-processing step applied to variables with the aim of scaling variables to a similar range.

The first data set ('direct marketing') analysed was obtained from one of South Africa's largest banks. The data set contains information about the bank's customers, the products they have with the bank, and their utilisation of and behaviour regarding those products. The target variable was binary: whether or not the customer responded to a direct marketing campaign for a personal loan. This data set contains 24 explanatory variables and 4720 observations.

The second data set ('protein structure') was obtained from the UCI Machine Learning Repository[22] and contains results of experiments performed by the Protein Structure Prediction Centre[23] on the latest protein structure prediction algorithms. These experiments were labelled the 'Critical Assessment of Protein Structure Prediction' experiments.[24] In computational biology, a persistent challenge is the prediction of tertiary structures of proteins.[25-29] Proteins assume three-dimensional tertiary structures and are therefore complex in nature. Structures are further influenced by a number of physico-chemical properties which further complicates the task of accurate prediction.[30] Protein structure prediction algorithms are algorithms that attempt to predict the tertiary structure of proteins.[26] These prediction algorithms have been refined over a number of years[31-34], but will still deviate when compared to samples of actual, experimentally determined structures. One way of measuring such deviations is through the root-mean-square-deviation.[26,27,35] Note that the protein structure prediction algorithms are in no way related to predictive modelling as defined in this paper, as they are specific to the field of protein assessment. The protein structure data set contains various physico-chemical properties of proteins, and the target variable is based on the root-mean-square-deviation measurement, indicating how much the predicted protein structures deviate from experimentally determined structures. The binary target used was whether or not the root-mean-square-deviation had exceeded a certain value (7.5). Our goal was therefore to determine what physico-chemical properties cause protein structure prediction algorithms to deviate more than the norm from experimentally determined protein structures. This data set contains nine explanatory variables and 45 730 observations.

The third data set ('credit application') was obtained from the Kaggle website (www.kaggle.com).[36] The data set is publicly available, and was used in a competition called 'Give me some credit', which ran from September to December 2011. The data set contains 10 characteristics of customers who applied for credit, and the target variable is binary, indicating whether or not the customer experienced a 90-day or longer delinquency. The data set is used in a number of studies covering various areas of predictive modelling.[37-40] All missing values (indicated by a 'NA' value) were substituted with a value of zero.

The fourth data set ('wine quality') was obtained from the UCI Machine Learning Repository.[22] The data comprise physico-chemical properties of wines that are extracted through analytical tests that can be easily performed on most wines. The data set was collected between May 2004 and February 2007.[41] The target variable was derived from a score between 0 and 10 which indicates the quality of the wine as scored by tasting experts. The binary target variable used for this analysis was whether or not the score was greater than 6, thereby indicating a great quality wine (only 20% of the wines scored greater than 6). The repository consisted of two data sets – one for white wines and one for red wines. For the purposes of this exercise, the two data sets were combined. The data set has 11 explanatory variables and 6497 observations.

The fifth data set ('chess king-rook vs king') is based on game theory and was obtained from the UCI Machine Learning Repository.[22] The data set is an 'Endgame database', which is a table of stored game theoretic values for the legal positions of the pieces on a chessboard. In this endgame, first described by Clarke[42], the white player has both its king and its rook left, whilst the black player only has its king left – it is widely known as the 'KRK endgame' and is still the focus of many studies[43-45]. The database stores the positions of each piece as well as the number of moves taken to finish the game from those positions assuming minimax-optimal play (black to move first). The target variable is binary, and indicates whether the game will be completed within 12 moves or less. Minimax-optimal play is an algorithm often used by computers to obtain the best combination of moves in a chess game and is based on the minimax game theory introduced by Neumann[46]. More information on this can be found in a number of texts, for example see Casti and Casti[47] and Russell and Norvig[48]. To the 6 explanatory variables another 12 derived variables were added (row distances, column distances, total distances and diagonal indicators). This data set contains 28 056 observations.

Table 1: Eight modelling techniques

Linear/non-linear	Modelling technique	Segmentation method used	Detailed description of modelling technique
Linear modelling technique	Logistic regression	Unsupervised	Unsupervised segmentation (k-means) with logistic regression
		Semi-supervised	Semi-supervised segmentation (SSSKMIV) with logistic regression
		Supervised	Supervised segmentation (decision trees) with logistic regression
Non-linear modelling techniques	Neural networks	No segmentation	Neural networks (AutoNeural Node in SAS Enterprise Miner)
	Support vector machines		Support vector machines (SVM node in SAS Enterprise Miner)
	Memory-based reasoning		Memory-based reasoning (MBR node in SAS Enterprise Miner)
	Decision trees		Decision trees (Decision Tree node in SAS Enterprise Miner)
	Gradient boosting		Gradient boosting (Gradient Boosting node in SAS Enterprise Miner)

The sixth data set ('insurance claim'), also obtained from the Kaggle website[36], contains information about bodily injury liability insurance. The competition was named 'Claim Prediction Challenge (Allstate)' and concluded in 2011. The binary target was whether or not a claim payment was made. The independent variables have been hidden, but according to the website, the data set contains information about the vehicle to which the insurance applies as well as some particulars about the policy itself. The data set itself has many observations (7.75 million), but events are rare (probability of occurrence around 1%). In order to reduce unnecessary computation time, the data set was therefore oversampled, which increased the event rate to around 33% (total observations on 14 782) with 12 explanatory variables. Oversampling in cases in which events are rare is a common technique applied in the industry.[49-51]

Results

Eight modelling techniques were compared using all six data sets. We compared the model performance achieved on linear modelling techniques (when first segmenting the data) to the accuracy of popular non-linear modelling techniques.

Table 2 summarises the performance of the modelling techniques when applied to the 'direct marketing' data set (as measured by the Gini coefficient calculated on the validation set). The gradient boosting technique achieved the best result on this data set, with decision tree segmentation running a close second. Neural networks could not converge to a model without overfitting, and the resulting Gini on the validation set is therefore effectively equal to zero. What can be seen additionally from Table 2 is that segmentation-based techniques take in positions two through to four as ranked by the Gini coefficient on the validation set.

Table 2: Direct marketing data set: Comparison of performance

Modelling technique	Best Gini obtained	Rank
Unsupervised segmentation (k-means) with logistic regression	27.11%	4
Semi-supervised segmentation (SSSKMIV) with logistic regression	27.89%	3
Supervised segmentation (decision trees) with logistic regression	33.70%	2
Neural networks	0%	8
Support vector machines	24.46%	5
Memory-based reasoning	21.95%	7
Decision trees	22.94%	6
Gradient boosting	35.31%	1

Table 3 summarises the Gini results of the various techniques as applied to the data set on 'protein tertiary structures'. As evidenced by the table, the ranking order of the techniques is completely different from the order seen in Table 2. As a start, gradient boosting ranks third from the bottom, at number six. The technique that achieves the best results in this case is memory-based reasoning. In Table 2, memory-based reasoning was ranked at position seven. The best-ranked segmentation-based technique for this data set is SSSKMIV in position two.

Table 4 shows that, for the 'credit application' data set, neural networks outperform all other techniques. In Tables 2 and 3, neural networks ranked last each time. However, in this case the structure of the data set evidently suited the technique well.

Similar to what was seen in Table 2, segmentation-based techniques take up positions two to four for this data set, with supervised segmentation (decision trees) performing best. At this point, a trend is emerging that segmentation-based techniques may not always render the best results, but seem to deliver results that are consistently amongst the top.

Table 3: Protein tertiary structures data set: Comparison of performance

Modelling technique	Best Gini obtained	Rank
Unsupervised segmentation (k-means) with logistic regression	66.88%	4
Semi-supervised segmentation (SSSKMIV) with logistic regression	70.37%	2
Supervised segmentation (decision trees) with logistic regression	66.43%	5
Neural networks	47.32%	8
Support vector machines	57.04%	7
Memory-based reasoning	80.33%	1
Decision trees	69.17%	3
Gradient boosting	57.89%	6

Table 4: Credit application data set: Comparison of performance

Modelling technique	Best Gini obtained	Rank
Unsupervised segmentation (k-means) with logistic regression	63.11%	4
Semi-supervised segmentation (SSSKMIV) with logistic regression	66.25%	3
Supervised segmentation (decision trees) with logistic regression	70.89%	2
Neural networks	72.20%	1
Support vector machines	31.47%	8
Memory-based reasoning	43.80%	7
Decision trees	48.41%	6
Gradient boosting	53.96%	5

Table 5 shows that for the 'wine quality' data set, segmentation-based techniques occupy the top two positions, with supervised segmentation (decision trees) in position four. The results are generally very close, with only decision trees and support vector machines not doing particularly well.

Table 6 shows that decision trees are best suited for the non-linear nature of the chess king-rook vs. king data set. This data set is the first for which segmentation-based techniques fail to be among the top two techniques, with supervised segmentation (decision trees) in third place.

Table 7 shows the results of the last data set to be analysed – the 'insurance claim prediction' data set. It can be seen from the table that the first two positions are again held by segmentation-based techniques, with SSSKMIV achieving the best results. The best non-segmentation-based technique is gradient boosting in position three followed by unsupervised k-means segmentation. The Gini coefficients for this application are low, so the relative difference between the 15.18% obtained by SSSKMIV and the 12.92% of gradient boosting is quite high.

Table 5: Wine quality data set: Comparison of performance

Modelling technique	Best Gini obtained	Rank
Unsupervised segmentation (k-means) with logistic regression	67.21%	1
Semi-supervised segmentation (SSSKMIV) with logistic regression	66.97%	2
Supervised segmentation (decision trees) with logistic regression	66.50%	4
Neural networks	66.64%	3
Support vector machines	59.66%	8
Memory-based reasoning	66.10%	5
Decision trees	60.86%	7
Gradient boosting	63.34%	6

Table 6: Chess king-rook vs. king data set: Comparison of performance

Modelling technique	Best Gini obtained	Rank
Unsupervised segmentation (k-means) with logistic regression	86.95%	5
Semi-supervised segmentation (SSSKMIV) with logistic regression	86.60%	6
Supervised segmentation (decision trees) with logistic regression	88.34%	3
Neural networks	25.47%	8
Support vector machines	74.81%	7
Memory-based reasoning	90.63%	2
Decision trees	93.34%	1
Gradient boosting	87.25%	4

Table 7: Insurance claim prediction data set: Comparison of performance

Modelling technique	Best Gini obtained	Rank
Unsupervised segmentation (k-means) with logistic regression	12.92%	4
Semi-supervised segmentation (SSSKMIV) with logistic regression	15.19%	1
Supervised segmentation (decision trees) with logistic regression	13.72%	2
Neural networks	10.22%	5
Support vector machines	10.06%	6
Memory-based reasoning	9.39%	7
Decision trees	8.69%	8
Gradient boosting	12.92%	3

Conclusions

Although it was not the focus of this paper to do an exhaustive comparison of modelling techniques, we provide an overview of how some of the more popular non-linear techniques perform when compared to segmented linear regression. Perhaps because of the diverse nature of the data sets used in this paper, it was interesting to see that no single technique dominated the top position. The Gini coefficients on the validation set of eight modelling techniques were compared. Specifically when considering the data from a local South African bank, gradient boosting performed the best. What was also clear was that the three segmentation-based techniques explored always performed well on all six data sets, even though other techniques demonstrated some significant inconsistency. Table 8 summarises the best performing technique for each data set. In addition, the table also shows the position, or rank, of the best performing segmentation-based technique. The consistency is clear from the fact that these three segmentation-based techniques usually take either position one or two, with only a single third place.

Table 9 provides another view on the consistency of the segmentation-based techniques. The table provides the average rank of each technique (calculated over all six data sets). The table was sorted from lowest average rank to highest average rank. As expected, the segmentation-based techniques do very well, taking the first three positions. SSSKMIV is rated first with an average rank of 2.8.

Table 8: Summary of results of alternative techniques compared to segmentation-based technique

Data set	Best technique	Position of best segmentation-based technique
Direct marketing	Gradient boosting	2
Protein tertiary structures	Memory-based reasoning	2
Credit application data	Neural networks	2
Wine quality	Unsupervised segmentation (k-means) with logistic regression	1
Chess king-rook vs. king	Decision trees	3
Insurance claim prediction	Supervised segmentation (decision trees) with logistic regression	1

Table 9: Average ranking position of modelling techniques over all six data sets

Modelling technique	Average rank
Semi-supervised segmentation (SSSKMIV) with logistic regression	2.8
Supervised segmentation (decision trees) with logistic regression	3.0
Unsupervised segmentation (k-means) with logistic regression	3.7
Gradient boosting	4.2
Memory-based reasoning	4.8
Decision trees	5.2
Neural networks	5.5
Support vector machines	6.8

We conclude that the SSSKMMIV algorithm (semi-supervised segmentation method), although not always outperforming unsupervised and supervised methods, can be a valuable tool to improve segmentation for predictive linear modelling, and does in many cases provide better segmentation than the traditional segmentation methods. The benefit of segmentation was also clearly illustrated in the six data sets used. We showed that the use of non-linear models might not be necessary to increase model performance when data sets are first segmented.

Acknowledgements

This work is based on research supported in part by the Department of Science and Technology (DST) of South Africa. The grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by DST-supported research are those of the authors and that the DST accepts no liability whatsoever in this regard.

Authors' contributions

D.G.B. was responsible for conceptualisation; methodology; data collection; data analysis; validation; data curation; writing revisions; and project leadership. T.V. was responsible for conceptualisation; sample analysis; data analysis; writing the initial draft; revisions; student supervision; and project management.

References

1. Hand DJ. What you get is what you want? – Some dangers of black box data mining. In: M2005 Conference Proceedings. Cary, NC: SAS Institute Inc.; 2005.
2. Baesens B, Roesch D, Scheule H. Credit risk analytics: Measurement Techniques, Applications, and Examples in SAS. New Jersey: Wiley; 2016.
3. Tevet D. Exploring model lift: Is your model worth implementing? Actuarial Rev. 2013;40(2):10–13.
4. Anderson R. The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation. New York: Oxford University Press; 2007.
5. Siddiqi N. Credit risk scorecards. Hoboken, NJ: John Wiley & Sons; 2006.
6. Thomas LC. Consumer credit models. New York: Oxford University Press; 2009. http://dx.doi.org/10.1093/acprof:oso/9780199232130.001.1
7. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. Berlin: Springer; 2001. https://doi.org/10.1007/978-0-387-21606-5_14
8. Hand DJ. Construction and assessment of classification rules. West Sussex: John Wiley & Sons; 1997.
9. Breed DG. Semi-supervised segmentation within a predictive modelling context. Potchefstroom: North-West University: 2017.
10. SAS Institute Inc. Predictive modelling using logistic regression (SAS Institute course notes). Cary, NC: SAS Institute Inc.; 2010.
11. Cross G. Understanding your customer: Segmentation techniques for gaining customer insight and predicting risk in the telecom industry. Paper 154-2008. Paper presented at: SAS Global Forum 2008. Available from: http://www2.sas.com/proceedings/forum2008/154-2008.pdf
12. Fico. Using segmented models for better decisions [document on the Internet]. c2014 [cited 2015 Jan 05]. Available from: http://www.fico.com/en/node/8140?file=9737
13. Breed DG, De La Rey T, Terblanche SE. The use of different clustering algorithms and distortion functions in semi supervised segmentation. In: Proceedings of the 42nd Operations Research Society of South Africa Annual Conference; 2013 September 15–18; Stellenbosch, South Africa. Available from: http://www.orssa.org.za/wiki/uploads/Conf/ORSSA2013_Proceedings.pdf
14. SAS Institute Inc. Applied analytics using SAS Enterprise Miner (SAS Institute Course Notes). Cary, NC: SAS Institute Inc.; 2015.
15. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. B Math Biophys. 1943;5(4):115–133. http://dx.doi.org/10.1007/BF02478259
16. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–297.
17. Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press; 2000. http://dx.doi.org/10.1017/CBO9780511801389
18. Li L. Support vector machines. In: Selected applications of convex optimization. New York: Springer; 2015. p. 17–52. http://dx.doi.org/10.1007/978-3-662-46356-7_2
19. Meyer D, Wien FHT. Support vector machines. Technical report. Boston: R Foundation for Statistical Computing; 2014.
20. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Statist. 2001;29(5):1189–1232. http://dx.doi.org/10.1214/aos/1013203451
21. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. http://dx.doi.org/10.1023/A:1010933404324
22. Lichman M. UCI Machine Learning Repository datasets [data sets on the Internet]. c2013 [cited 2016 May 06]. Available from: http://archive.ics.uci.edu/ml.
23. Protein Structure Prediction Center [homepage on the Internet]. c2015 [cited 2016 Jun 04]. Available from: http://predictioncenter.org/.
24. Kryshtafovych A, Monastyrskyy B, Fidelis K. CASP prediction center infrastructure and evaluation measures in CASP10 and CASP ROLL. Proteins: Struct Funct Bioinf. 2014;82(Suppl 2):7–13.
25. Fraenkel A. Complexity of protein folding. Bull Math Biol. 1993;55(6):1199–1210. http://dx.doi.org/10.1007/BF02460704
26. Mishra A, Rana PS, Mittal A, Jayaram B. D2N: Distance to the native. BBA-Proteins Proteom. 2014;1844(10):1798–1807.
27. Rana PS, Sharma H, Bhattacharya M, Shukla A. Quality assessment of modeled protein structure using physicochemical properties. J Bioinform Comput Biol. 2015;13(2), Art. #1550005, 19 pages. http://dx.doi.org/10.1142/S0219720015500055
28. Searls D. Grand challenges in computational biology. In: Salzberg S, Searls D, Kasif S, editors. Computational methods in molecular biology. Amsterdam: Elsevier; 1998. p. 3–10.
29. Unger R, Moult J. Finding the lowest free energy conformation of a protein is an NP-hard problem: Proof and implications. Bull Math Biol. 1993;55(6):1183–1198. http://dx.doi.org/10.1007/BF02460703
30. Anfinsen CB. Principles that govern the folding of protein chains. Science. 1973;181(4096):223–230. http://dx.doi.org/10.1126/science.181.4096.223
31. Dhingra P, Jayaram B. A homology/ab initio hybrid algorithm for sampling near-native protein conformations. J Comput Chem. 2013;34(22):1925–1936. http://dx.doi.org/10.1002/jcc.23339
32. Jayaram B, Dhingra P, Lakhani B, Shekhar S. Targeting the near impossible: Pushing the frontiers of atomic models for protein tertiary structure prediction. J Chem Sci. 2012;124(1):83–91. http://dx.doi.org/10.1007/s12039-011-0189-x
33. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004;32(2):W526–W531. http://dx.doi.org/10.1093/nar/gkh468
34. Lambert C, Léonard N, De Bolle X, Depiereux E. ESyPred3D: Prediction of proteins 3D structures. Bioinformatics. 2002;18(9):1250–1256. http://dx.doi.org/10.1093/bioinformatics/18.9.1250
35. Cozzetto D, Kryshtafovych A, Fidelis K, Moult J, Rost B, Tramontano A. Evaluation of template-based models in CASP8 with standard measures. Proteins. 2009;77(S9):18–28. http://dx.doi.org/10.1002/prot.22561
36. Kaggle [homepage on the Internet]. c2016 [cited 2016 Sep 23]. Available from: http://www.kaggle.com.
37. Bahnsen AC, Aouada D, Ottersten B. Example-dependent cost-sensitive logistic regression for credit scoring. In: Proceedings of the 13th International Conference on Machine Learning and Applications (ICMLA); 2014 December 3–5; Detroit, MI, USA. Available from: https://doi.org/10.1109/ICMLA.2014.48

38. Sharma D. Elements of Optimal Predictive Modeling Success in Data Science: An Analysis of Survey Data for the 'Give Me Some Credit' Competition Hosted on Kaggle. Available at SSRN 2227333. 2013. https://doi.org/10.2139/ssrn.2227333

39. Sitar M, Rašeta J, Klešćek A. Implementation of data mining techniques in credit scoring. In: Marković A, Rakočević SB, editors. Proceedings of the XIV International Symposium Symorg 2014: New business models and sustainable competitiveness; 2014 June 6–10; Zlatibor, Serbia. FON; 2014. p. 130.

40. Zhou L, Wang H. Loan default prediction on large imbalanced data using random forests. TELKOMNIKA Indones J Electr Eng. 2012;10(6):1519–1525. http://dx.doi.org/10.11591/telkomnika.v10i6.1323

41. Cortez P, Cerdeira A, Almeida F, Matos T, Reis J. Modeling wine preferences by data mining from physicochemical properties. Decis Support Syst. 2009;47(4):547–553. http://dx.doi.org/10.1016/j.dss.2009.05.016

42. Clarke M. A quantitative study of king and pawn against king. In: Clarke MRB, editor. Advances in computer chess. Edinburgh: Edinburgh University Press; 1977. p. 108–118.

43. Bain M. Experiments in non-monotonic learning. Paper presented at: The Eighth International Workshop on Machine Learning; 2014 September 6–7; Strasbourg, France. p. 380–384.

44. Bramer MA. Machine-aided refinement of correct strategies for the endgame in chess. Adv Comp Chess. 2014;3:93–112.

45. Cohen WW. Compiling prior knowledge into an explicit bias. In: Proceedings of the Ninth International Conference on Machine Learning. Burlington, MA: Morgan Kaufmann; 1992. p. 102–110.

46. Neumann Jv. Zur theorie der gesellschaftsspiele [On the theory of social games]. Mathematische Annalen. 1928;100(1):295–320. German. http://dx.doi.org/10.1007/BF01448847

47. Casti JL, Casti JL. Five golden rules: Great theories of 20th-century mathematics and why they matter. New York: John Wiley & Sons; 1996.

48. Russell SJ, Norvig P. Artificial Intelligence: A modern approach. 2nd ed. New Delhi: Pearson Education; 2003.

49. Chang CY, Hsu MT, Esposito EX, Tseng YJ. Oversampling to overcome overfitting: Exploring the relationship between data set composition, molecular descriptors, and predictive modeling methods. J Chem Inform Modeling. 2013;53:958–971. https://doi.org/10.1021/ci4000536

50. Taft LM, Evans RS, Shyu CR, Egger MJ, Chawla N, Mitchell JA, et al. Countering imbalanced datasets to improve adverse drug event predictive models in labor and delivery. J Biomed Inform. 2009;42:356–364. https://doi.org/10.1016/j.jbi.2008.09.001

51. Yap BW, Rani KA, Rahman HAA, Fong S, Khairudin Z, Abdullah NN. An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In: Herawan T, Deris M, Abawajy J, editors. Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013); Lecture Notes in Electrical Engineering, vol 285. Singapore: Springer; 2014. p. 13–22. https://doi.org/10.1007/978-981-4585-18-7_2

The potential of South African timber products to reduce the environmental impact of buildings

AUTHORS:
Philip L. Crafford[1]
Melanie Blumentritt[1]
C. Brand Wessels[1]

AFFILIATION:
[1]Department of Forest and Wood Science, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Philip Crafford

EMAIL:
pcrafford@sun.ac.za

DATES:
Received: 16 Nov. 2016
Revised: 25 Apr. 2017
Accepted: 17 May 2017

KEYWORDS:
life-cycle assessment; pine; steel; novel truss materials; green building

HOW TO CITE:
Crafford PL, Blumentritt M, Wessels CB. The potential of South African timber products to reduce the environmental impact of buildings. S Afr J Sci. 2017;113(9/10), Art. #2016-0354, 8 pages. http://dx.doi.org/10.17159/sajs.2017/20160354

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Hans Merensky Foundation

South Africa was the first country in Africa to implement a locally developed green building rating tool and has a growing number of rated green building projects. The method of life-cycle assessment can help to compare and assess the environmental performance of building products. At present, more than 70% of all sawn timber in South Africa is used in buildings, mainly in roof structures. Light gauge steel trusses have recently also been gaining market share. However, to date, no studies have been conducted that quantify and compare the environmental impacts of the different roof truss systems in South Africa. We thus compared several roof truss systems (South African pine, Biligom and light gauge steel) found in low- and medium-income house designs in South Africa using a simplified life-cycle assessment approach. Our results show that the two timber systems had overall the lowest environmental impact. Although the difference between the timber systems was small, light gauge steel had a 40% higher normalised impact over all assessed environmental impact categories. The benefit of biogenic carbon dioxide present in timber proved to play a significant positive role in the global warming potential impact and could even be further reduced if wood were used to generate energy at its end-of-life. This study demonstrates the potential advantage of using local timber products to reduce the environmental impact of the truss and building industry in South Africa.

Significance:

- Timber truss systems showed overall lower environmental impact than light gauge steel trusses, with implications for green building.

Introduction

Buildings are major emitters of carbon dioxide and contribute significantly to global climate change.[1,2] A growing global awareness of the environmental footprint of buildings and the necessity to lower greenhouse gas emissions has led to the implementation of green building practices and the introduction of green building rating tools that have been used to measure the environmental impact and sustainability of buildings since the 1990s.[3] Numerous studies have shown that substituting steel, concrete and brick materials with renewable and sustainable wood products can significantly lower the environmental impact of a building over its lifetime.[4-9]

Residential roof truss construction in South Africa is the single biggest user of locally produced structural timber. According to Crickmay and Associates[10], more than 70% of all structural timber is used in the local building market. Structural timber in South Africa is mostly South African pine (various *Pinus* species). In addition eucalyptus (mostly *Eucalyptus grandis*) timber is also used in structural applications, such as laminated beams and Biligom – a new, moist, glued, finger-jointed structural timber product for truss systems.[11] Traditionally, structural steel is known for its ability to cover large spans and use in high stress applications such as reinforced concrete. Recently, light gauge steel (LGS) construction as well as LGS roof truss systems have gained a noticeable market share and offer another option as roof truss material. With steel prices currently low, many building projects and smaller roof spans with steel have become economically viable options in South Africa, and in many cases replaced wood as the preferred truss material.[12,13]

In a combined life-cycle assessment (LCA) and cost study performed by Worth et al.[14], in which softwood timber trusses were compared with imported LGS in New Zealand, the authors found that LGS requires at least 6.65 times more energy to manufacture than wood. In a study by Bolin and Smith[15], it was found that in their manufacture, use and disposal, CCA-treated wood guard rails require lower fossil fuel use, produce lower greenhouse gas (GHG) emissions and have lower environmental impacts in the acidification, smog potential and eco-toxicity categories compared with that of galvanised steel posts. Ximenes and Grant[8] assessed the GHG benefits of the use of wood products compared with those of steel-reinforced concrete in two popular house designs in Sydney, Australia. The timber frame option for the roof resulted in a net GHG emission reduction ranging from 51% to 66% compared with steel frames for the equivalent roofing material.

Governments, architects, developers and the general public are under an increasing obligation to make environmentally responsible decisions when it comes to selecting building materials and methods.[16] South Africa was the first country in Africa to implement a locally developed green building rating tool and has a growing number of rated green building projects.[17] At the same time, however, marketing is used to promote materials and buildings as green and environmentally sound without concrete evidence in support of these claims.

Nearly 70% of all sawn timber in South Africa is utilised in buildings, specifically in roof trusses. LGS trusses are also gaining market share. However, to date no studies have been conducted quantifying the environmental impacts of the different truss systems in South Africa. End-users of trusses, therefore, do not have the necessary information to make environmentally responsible choices when selecting a truss system. Additionally, manufacturers of both timber and LGS trusses have little information to guide them in reducing the environmental impacts of their processes and products.

In this study, we investigated and compared the potential environmental impact of different roof truss systems typically found in low- and medium-income house designs in South Africa using a simplified LCA approach. Environmental impacts were compared over 11 different impact categories. We present the potential environmental impact of the modelled products and discuss adjustments and assumptions made with regard to the availability of South Africa specific life-cycle inventory (LCI) data and validity of obtained results.

Life-cycle assessment is a methodical framework for estimating and assessing the potential environmental impacts of a product system or process over its entire life cycle, including raw material extraction, manufacturing, use, and end-of-life disposal and/or recycling.[18] Thus, LCA is often considered a 'cradle-to-grave' approach to evaluate environmental impacts.[19] The International Organization for Standardization (ISO) adopted an environmental management standard in the 1990s as part of its 14000 standards series, with the 14040 series focusing on establishing methodologies for LCA.[20,21] The ISO standards define a four-stage interactive framework for conducting LCA analysis. The first stage is the definition of the goal and scope of the study including the establishment of the functional unit, system boundaries and quality criteria for LCI data. Life-cycle inventory, the second stage, deals with the collection and synthesis of information of system inputs and outputs of material and energy flows and associated environmental impacts in all stages of the life cycle. During the life-cycle impact assessment (LCIA), the third stage, these environmental impacts are assigned to different environmental impact categories and by means of characterisation factors, the contribution of each constituent is calculated for different environmental impact categories (e.g. global warming potential, human toxicity, acidification, resource depletion, land use). The final stage is the interpretation of the results from both LCI and LCIA.[20,21]

Goal and scope

Objective

The goal of this study was to assess the potential environmental impact associated with the manufacture, use and disposal of timber and light gauge steel roof truss systems commonly found and used in South Africa. We compared three different truss materials – South African pine (S5), LGS and Biligom – in two house designs (Table 1). Biligom is a new sawn timber product made from green finger-jointed *E. grandis* wood. A 42-m^2 Reconstruction and Development Programme (RDP) house and a 168-m^2 single-story family house were chosen to represent commonly found house sizes in the South African lower- and middle-income market. Concrete tiles were selected as the roof cover material.

Table 1: Experimental design summary

Alternative	Truss material	Number of trusses	Cover material	House footprint
1	SA pine S5	10	Concrete tiles	42 m^2 (6x7 m)
2	Biligom	10	Concrete tiles	42 m^2 (6x7 m)
3	Light gauge steel	7	Concrete tiles	42 m^2 (6x7 m)
4	SA pine S5	16	Concrete tiles	168 m^2 (14x12 m)
5	Biligom	16	Concrete tiles	168 m^2 (14x12 m)
6	Light gauge steel	12	Concrete tiles	168 m^2 (14x12 m)

Limitations

A significant portion of the overall life-cycle energy requirements of buildings is from occupational energy use. However, owing to time and data constraints, occupational energy consumption over the design life of the roof and associated building was not considered in this study. In reality, different roofing materials will have, next to their own embodied energy, an impact on the energy usage required for space heating and ventilation and further investigation is necessary to address this issue properly. The assessment of the roof configurations is limited to the environmental factors associated with each type of roof truss system, excluding the cover material (i.e. concrete tiles) and the supporting building structure. Costing was also not included in the analyses.

Methodology

A detailed description of the LCA methodology and framework is available in the ISO 14040 Environmental Management series.[20,21] Many of the recommendations set out in these documents are above and beyond the scope of the current study; however, the sections of these guidelines relevant to this study were followed.

The functional unit, as defined in ISO 14041, was chosen for this study as the quantity of materials required to construct the roof truss system of a house with a predefined footprint (i.e. 42 m^2 or 168 m^2). Both theoretical house designs have cement block walls. All structural components required that make up the roof structure were considered (namely truss material, bracing material, battens, purlins, nails and screws). The cover material (i.e. concrete tiles) and insulation material were not included, but were considered for the design (e.g. in terms of load-bearing capacity of the roof structure). The roofs were designed with a 17.5° pitch and for a 50-year service life in the Western Cape Province of South Africa. The roof structures were calculated and designed by MiTek South Africa (Pty) Ltd engineers (Cape Town) according to national timber construction standards. MiTek design software provided a detailed material and cutting list for all structural components per design, either per mass or per volume (Table 2). Waste produced from cutting standard lengths to size was not accounted for. We assumed that no maintenance work or replacements would be necessary over the design lifespan.

Biligom structural timber is 25–35% stronger than South African pine structural grade S5 in terms of flexural properties, i.e. bending strength and stiffness[11]; in this theoretical comparison, because of current design constraints and data availability, Biligom was assumed to be equal in volume/dimensions to South African pine (S5).

Table 2: Roof truss systems with the mass and volume per material category for each alternative

Alternative	SA pine (S5)		Biligom		Light gauge steel
	m^3	kg	m^3	kg	kg
1	1.33	598.5			22.5
2			1.33	798.0	23.5
3					167.8
4	6.05	2722.5			180.6
5			6.05	3630.0	186.6
6					1094.0

Wood density is taken as air dry density for South African pine (450 kg/m^3) and partially wet density for Biligom (600 kg/m^3).

Both high strength (ISQ 550-3T) and low strength (ISQ 300) components are used in MiTek truss systems. The steel is similar in production and treatment across the entire manufacturing process. Here it is assumed that the same type of steel is used for all components. All light gauge steel material is galvanised at 200 g/m^2.

End-plates are used as part of the Biligom product at 0.96 kg/m^3 Biligom and both timber systems make use of nail plates as truss component connectors.

Life-cycle inventory

In this study, openLCA 1.4.2 modelling software was used to determine the LCI. The materials used in the LCIs were assumed to be sourced and processed locally. As there is little to no LCI data available for South Africa, global data sets from the ecoinvent database 3.1 (Weidema et al.[22])

were used. Adjustments were made to existing processes in the ecoinvent database when possible to better represent local conditions (e.g. by using local electricity data available in ecoinvent or adjusting conversion factors).

Timber

We assessed two types of timber: South African pine in grade S5 and Biligom. Plantation forestry for pine and eucalyptus is practised In South Africa. LCI data from the Australian life-cycle inventory database (AUSLCI) was used and integrated into the ecoinvent database to model the softwood forestry process, as it reasonably represents local conditions. Sawmilling, drying and planing of the timber were modelled using ecoinvent processes for softwood, but adjusted to use South African specific conversion factors and electricity.

Biligom is a recent development of finger-jointed moist glued eucalyptus timber and original LCI data were gathered from BILIGOM® International (Pty) Ltd. The AUSLCI process for hardwood (eucalyptus) forestry was used to model the forestry process. Both product systems were modelled in openLCA using the ecoinvent database for background data.

Depending on the region in South Africa, both pine and eucalyptus timber used in load-bearing applications need to be preservative treated to comply with national building codes. Biligom uses TanalithE as preservative and copper chromated arsenate (CCA) was chosen for pine, as it is widely used in South Africa. Original LCI data on the chemical composition of both preservatives used locally were provided by Arch Wood Protection (SA) Pty Ltd and modelled in openLCA using the ecoinvent database for background data on chemicals, preservative production and pressure treatment.

Light-gauge steel

Light-gauge steel is made from galvanised sheet material, on continuous zinc coating lines from either cold-rolled (thickness range of 0.27 mm to < 2.0 mm) or hot-rolled (thickness of 2.01–3.0 mm) steel in coil form. It is produced to the requirements of a range of national and international standards as well as Mittal Steel South Africa's ISQ standards.[23] Continuous zinc-coated cold-rolled sheet metal, also known as LGS, and the machining thereof was modelled based on rest-of-world steel data, available in the ecoinvent database 3.1, including processes for steel production, sheet rolling, metal working and zinc coating. The rest-of-world data are assumed to closer reflect local process conditions than are European or global data sets, especially in terms of the primary energy mix as it was not feasible to adjust all background processes included in LGS production to use South African electricity data.

Transportation

Transportation of materials to the processing facilities and from there to the building site in the Western Cape was included. We assumed that the LGS was sourced from the Gauteng Province, Biligom from the plant in Tzaneen, Limpopo and pine timber was standard averaged and originated in the Southern Cape and Limpopo Provinces. At the end-of-life, it was assumed that all materials were transported over 50 km to their respective final destination (e.g. for incineration, landfilling or recycling).

End-of-life

Formal recycling and burning of wood waste for energy was not considered as it is currently not common practice in South Africa. According to the South African Wood Preservers Association's treated timber guidelines, treated timber should be disposed of at a registered landfill site.[24] However, in South Africa, significant amounts of waste wood are used in peri-urban and rural areas as fuel for cooking and heating. A study performed by Niyobuhungira[25] showed that more than 50% of the residential fuel wood used in peri-urban areas in the Western Cape was CCA treated. In this study we chose disposal of timber by incineration, modelled with processes from the ecoinvent database as the most likely final fate scenario.

For the LGS, no recycling benefits were considered in the disposal phase as locally manufactured galvanised LGS is mainly produced from virgin material and the majority of steel scrap is exported and reused outside South Africa.[26]

Life-cycle impact assessment

All inputs and outputs considered in the cradle-to-grave analyses, and intermediate steps, were analysed in openLCA 1.4.2 with the CML baseline impact assessment method version 4.4 as of January 2015 (GreenDelta[27]) including normalisation data for different countries and years and using physical allocation. Additionally, impact category GWP_{100} was calculated without including biogenic carbon dioxide sequestration and emissions, thus assuming carbon neutrality of biogenic carbon dioxide.

Results and discussion

The potential environmental impact of the three roof truss assemblies was assessed and compared. Both cradle-to-gate and cradle-to-grave results are presented below. Table 3 shows the cradle-to-gate results of the 42-m^2 and 168-m^2 houses. Over all categories, Biligom has the lowest impact in most categories, closely followed by pine, and LGS has the highest impact. The difference between the two timber alternatives is small compared to the differences between them and LGS. The order of impact in the individual categories is the same for the larger truss assemblies. The impact in the individual categories is on average 4.5 times higher for the two timber alternatives and 6.5 times higher for LGS between the 42-m^2 and 168-m^2 house sizes. These differences are explained and directly correlated to the material volume ratio, required per material alternative as displayed in Table 2. It is interesting to note that although the timber alternatives use more trusses per house, the LGS system mass ratio is higher between the two house design footprints (Table 1).

Only the global warming potential (GWP) has negative values indicating a positive impact at the gate. More specifically, the results indicate the amount of carbon dioxide equivalents sequestrated in the material at this stage minus carbon dioxide emissions from processing and excluding emissions from end-of-life. Table 4 shows the same results as in Table 3 from cradle-to-grave. As expected, there is mostly a small increase in all categories and the timber alternatives are better than LGS. The most significant change can be seen in the GWP_{100}, which is a result of the inclusion of emissions from wood incineration at the end-of-life of the timber systems. A significant increase in fossil fuel depletion and eutrophication for the wood alternatives and aquatic ecotoxicity for LGS must also be attributed to the end-of-life treatment as well as transportation processes.

Pine showed significantly higher human toxicity impact values compared to the others because of the CCA treatment process. According to the LCA process contribution analysis, chromium oxide production is responsible for more than 90% of the human toxicity impact of pine from cradle-to-gate. The higher photochemical oxidation impact value for Biligom is again because of the carbon monoxide emissions created by the forest management process. The forest management LCI data used in the Biligom LCA (the best available data) are from an Australian-based hardwood management process which used natural gas as part of their energy mix, which was responsible for 88% of the photochemical oxidation impact.

Over the last decade, carbon sequestration, carbon footprints and carbon emissions have become globally familiar terms. GWP is often one of the key impact factors when assessing the environmental performance of building materials. Timber is unique in the sense that trees sequestrate carbon dioxide during growth. By using wood in long-lived products, the re-emission can be delayed; additionally, by using wood products and by-products for energy generation, emission associated with fossil fuels can be avoided. Furthermore, wood products generally require less energy for manufacturing than equivalent alternatives.[7,28-30] There is an ongoing debate in the research community on how to treat biogenetic carbon emissions.[31,32] While the assumption of carbon neutrality is true given a long time perspective, climate neutrality is a different matter.

In order to better understand the climate change impact of using wood compared to LGS in this study, Figures 1 to 3 present a more differentiated view of the GWP and associated carbon dioxide streams.

Figure 1 shows the cradle-to-grave GWP incline for the three materials and the two house sizes. The graph clearly indicates that the two timber alternatives follow a similar near-flat GWP impact trend, whereas the LGS system shows a sharp increase between the small and bigger house footprints. Once again, this increase can be explained by the higher material mass ratio required to scale up the LGS systems from the 42-m^2 to the 168-m^2 house, compared to the timber alternatives. Note that because only two house footprints were analysed, the gradients in this graph are not equitable, but rather show a trend.

The rest of the analyses will focus on the 42-m^2 house roof designs.

Global warming potential is expressed in kilograms carbon dioxide equivalents (kg CO_2 eq.) and represents the impact of a number of gases (e.g. carbon monoxide, carbon dioxide, methane, HFC) standardised with their lifespan in the atmosphere to a unit of carbon dioxide.

Table 3: Cradle-to-gate roof truss alternative impact assessment summary for the two roof designs

Impact category	42-m^2 house			168-m^2 house			Reference unit
	Pine (1)	Biligom (2)	Light gauge steel (3)	Pine (4)	Biligom (5)	Light gauge steel (6)	
Acidification potential	3.43	3.13	9.28	19.93	18.63	60.53	kg SO_2 eq.
GWP_{100}	-919	-1224	988	-3721	-5100	6445	kg CO_2 eq.
Depletion of abiotic resources – elements, ultimate reserves	0.04	0.02	0.11	0.22	0.14	0.74	kg antimony eq.
Depletion of abiotic resources – fossil fuels	3301	3229	8918	19175	18923	58145	MJ
Eutrophication	1.20	1.14	3.50	7.10	6.85	22.82	kg PO_4--- eq.
Freshwater aquatic ecotoxicity	268	233	1035	1706	1552	6751	kg 1,4-dichlorobenzene eq.
Human toxicity	8193	813	2640	38503	4956	17218	kg 1,4-dichlorobenzene eq.
Marine aquatic ecotoxicity	7.02E+05	5.87E+05	2.28E+06	4.26E+06	3.75E+06	1.49E+07	kg 1,4-dichlorobenzene eq.
Ozone layer depletion	3.61E-05	3.21E-05	5.84E-05	1.90E-04	1.70E-04	3.80E-04	kg CFC-11 eq.
Photochemical oxidation	0.26	0.95	0.43	1.37	4.53	2.77	kg ethylene eq.
Terrestrial ecotoxicity	18.68	10.97	69.26	117	82.77	451	kg 1,4-dichlorobenzene eq.

Table 4: Cradle-to-grave roof truss alternative impact assessment summary for the two roof designs

Impact category	42-m^2 house			168-m^2 house			Reference unit
	Pine (1)	Biligom (2)	Light gauge steel (3)	Pine (4)	Biligom (5)	Light gauge steel (6)	
Acidification potential	4.21	4.46	9.52	23.60	24.81	62.07	kg SO_2 eq.
GWP_{100}	85	164	1038	873	1242	6769	kg CO_2 eq.
Depletion of abiotic resources – elements, ultimate reserves	0.04	0.02	0.11	0.23	0.14	0.74	kg antimony eq.
Depletion of abiotic resources – fossil fuels	5237	6513	9556	28281	34165	62308	MJ
Eutrophication	1.59	1.72	3.97	9.08	9.72	25.85	kg PO_4--- eq.
Freshwater aquatic ecotoxicity	737	744	4344	5379	5447	28328	kg 1,4-dichlorobenzene eq.
Human toxicity	8284	967	2790	38983	5726	18191	kg 1,4-dichlorobenzene eq.
Marine aquatic ecotoxicity	8.88E+05	8.27E+05	3.32E+06	5.59E+06	5.34E+06	2.17E+07	kg 1,4-dichlorobenzene eq.
Ozone layer depletion	6.05E-05	7.25E-05	7.33E-05	3.10E-04	3.60E-04	4.80E-04	kg CFC-11 eq.
Photochemical oxidation	0.29	1.00	0.44	1.51	4.76	2.85	kg ethylene eq.
Terrestrial ecotoxicity	19.28	12.44	69.62	120	89.62	453	kg 1,4-dichlorobenzene eq.

Anthropogenic carbon dioxide emissions are produced from various sources, such as fossil fuel use, waste material decomposition and organic material burning. The carbon dioxide flows over the life cycle of South African pine and Biligom are displayed in Figure 2. Three major carbon dioxide flows were captured in both GWP data reports: sequestrated carbon dioxide from the air and biogenic and fossil-derived carbon dioxide emissions. According to the US Environmental Protection Agency:

> *Biogenic CO_2 emissions are defined as CO_2 emissions related to the natural carbon cycle, as well as those resulting from the production, harvest, combustion, digestion, fermentation, decomposition, and processing of biologically based materials.*[33]

The sequestrated carbon dioxide in the air is a negative value because of the carbon that is stored in the tree through photosynthesis during growth. The biogenic carbon dioxide emissions in Figure 2 are 99% attributed to the incineration process whereas the fossil-derived carbon dioxide emissions are mainly attributed to manufacturing and transport processes. The difference in the magnitude of the carbon dioxide flows between the two timber systems is interesting to note. The lower biogenic carbon dioxide levels for pine can be explained by the lower material density. The slightly lower fossil carbon dioxide level for pine is mostly as a result of the shorter transportation distance to the building site and also a lower density (smaller mass to transport). Fossil fuel impact breakdown per alternative from the manufacturing stage, transport and disposal can be seen in Figure 4 to accentuate the transportation impact.

In theory, adding sequestrated carbon dioxide from the air and the biogenic carbon dioxide emission should be close to a net result of zero. By analysing the flows for both materials visually, it is evident that these two carbon dioxide flows do not exactly match up, but show a slight negative carbon dioxide net result. The most likely explanation for this negative net result is a difference in wood volume in the forestry background data, compared to the wood used in the trusses and the wood used in the modelled, Swiss-based, incineration process. Furthermore, the incineration process does not emit all the carbon contained in the wood as pure carbon dioxide. Although timber sequesters carbon dioxide in the growing phase, by adding the three types of carbon dioxide flows as seen in Figure 2, both pine and Biligom still result in a small positive carbon dioxide footprint.

Therefore, under a general simplified assumption of carbon neutrality of biomass, a closer look at the GWP (excluding biogenic carbon monoxide, carbon dioxide and methane flows) can help in the understanding of the global warming impact of the truss alternatives (Figure 3). This time not considering carbon dioxide, the net GWP impact of the LGS truss system is only about double the two wood alternatives. Both wood alternatives have a large contribution attributed to transportation-associated emission from the factory to the building site. This finding highlights the importance of the transportation method and resource location. Although alternative transportation methods – i.e. shipping and rail – might be more environmentally friendly, it was not part of the scope of this study. The final stage (i.e. site to grave) includes incineration of all three truss systems and shows a non-significant overall non-biogenic impact contribution compared to the cradle-to-gate and cradle-to-site impact.

Figure 4 displays the fossil fuel depletion per life-cycle stage. A similar trend to the contribution profile for the non-biogenic GWP (Figure 3) can be seen, with a large contribution from transportation to the wood alternatives, especially for Biligom.

While GWP and fossil fuel depletion are important and relatively easy to understand impact factors, to assess the largely fossil fuel based climate change impact of building products, other environmental indicators need to be considered for a holistic evaluation of the potential environmental impact of building materials beyond GWP. In the following section, normalisation was used to evaluate the overall environmental impact between truss systems based on the 11 baseline impact categories. Normalisation is a simple technique to equate different categories and magnitudes by adjusting values measured on different scales to a notionally common scale. In Table 5, normalised indices of each cradle-to-grave impact category for all three truss systems are displayed. In each case, the LGS impact was set as one and the remaining two in relation to one. Finally, the combined or pooled normalised impact was computed by repeating the process using the total normalised values per truss system. Equal weighting was used to compute the compiled impact.

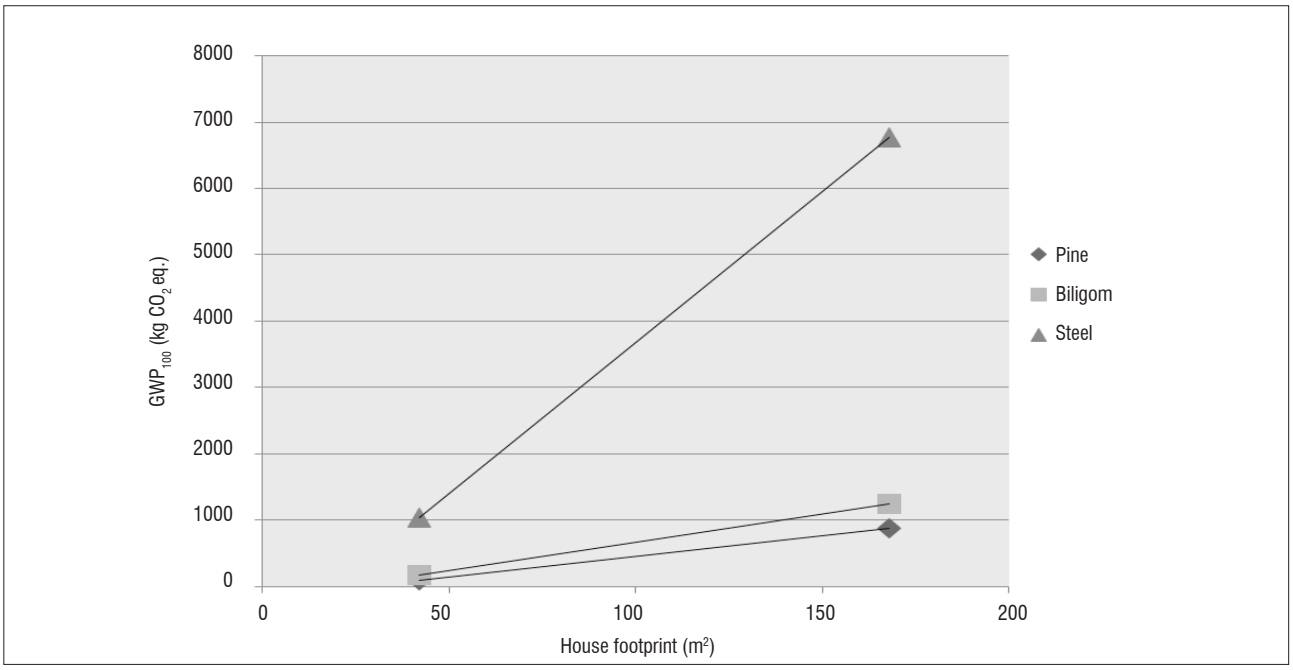


Figure 1: Global warming potential (GWP) gradient for South African pine, Biligom and light gauge steel for 42-m^2 and 168-m^2 houses.

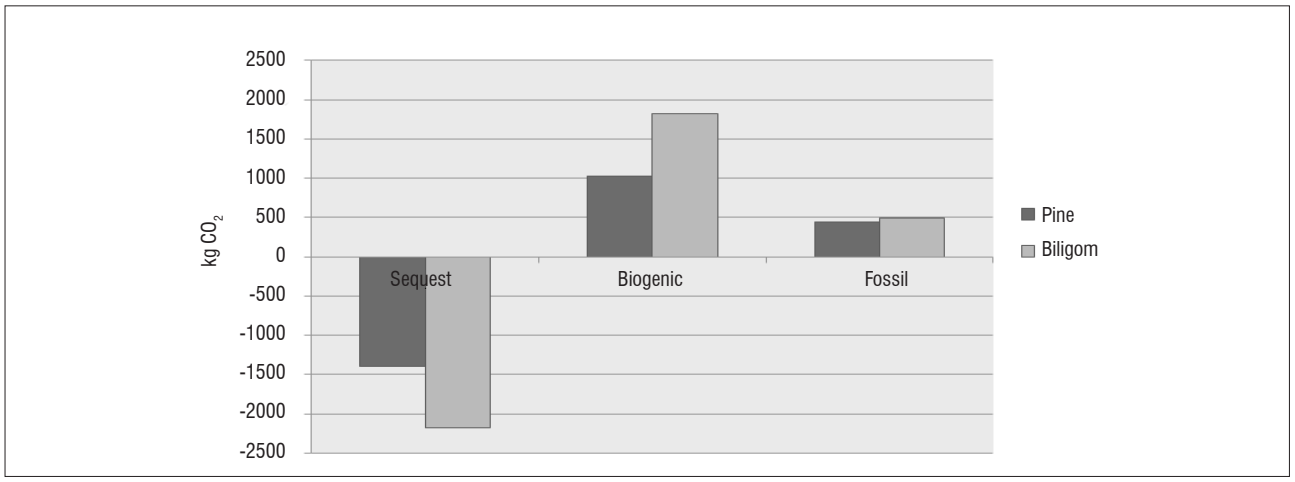


Figure 2: Carbon dioxide flow of South African pine and Biligom for the 42-m² roof design.

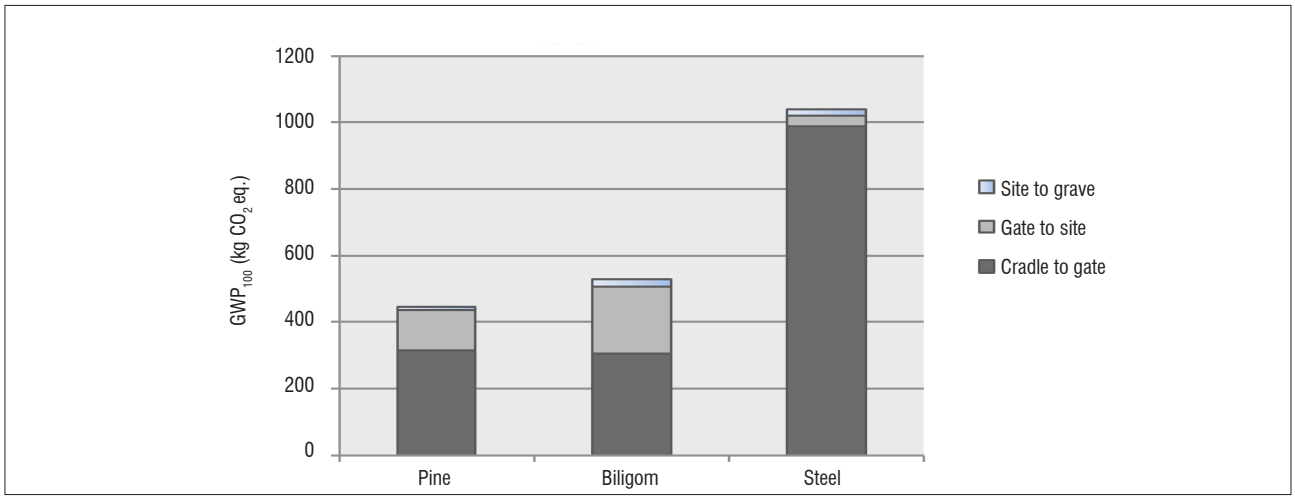


Figure 3: Global warming potential (GWP), excluding biogenic carbon monoxide, carbon dioxide and methane impact per life-cycle stage for the 42-m² roof design.

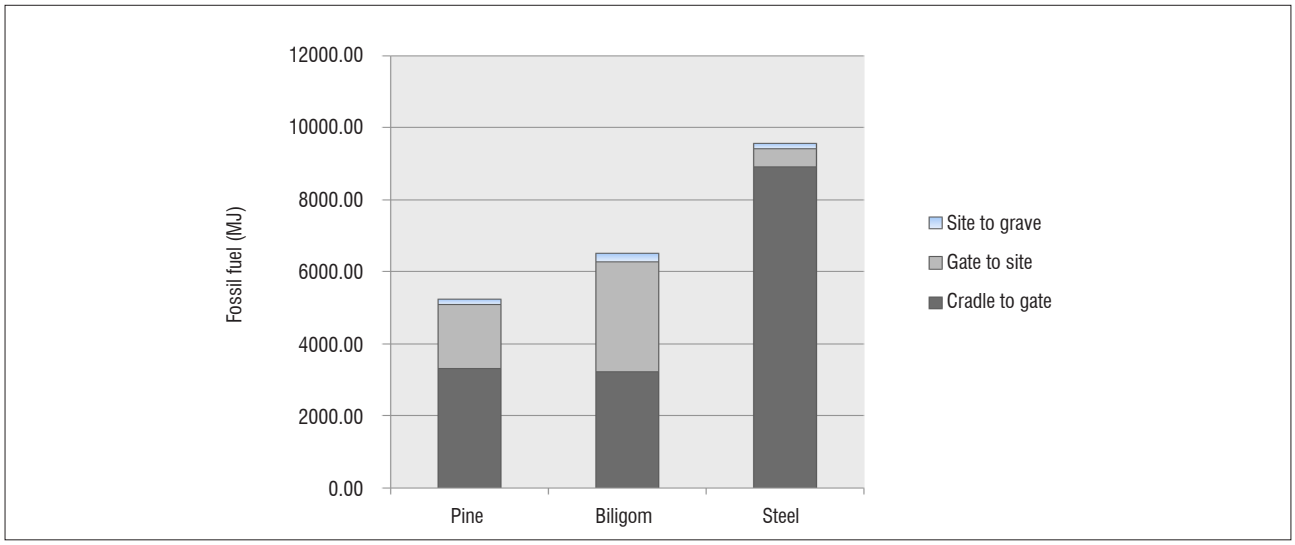


Figure 4: Depletion of abiotic resources/ fossil fuel (MJ) per life-cycle stage for the 42-m² roof design.

This method indicates that the overall environmental performance of the two timber systems is about 40% better than that of the LGS system. It also shows that one should be cautious of considering only one impact category to evaluate materials. For example, considering only climate change or human toxicity potential will portray a skewed picture. However, considering all impact data and results presented in this study, both timber truss systems outperform LGS but indicate a similar or higher impact in the human toxicity, ozone layer depletion and photochemical oxidation categories.

Table 5: Combined cradle-to-grave normalised impact per alternative material

Normalised impact for 42-m² roofs	Normalised indices		
Impact category	**Pine**	**Biligom**	**Steel**
Acidification potential – average Europe	0.44	0.47	1
Climate change – GWP_{100}	0.08	0.16	1
Depletion of abiotic resources – elements	0.34	0.17	1
Depletion of fossil fuels	0.55	0.68	1
Eutrophication – generic	0.40	0.43	1
Freshwater aquatic ecotoxicity – FAETP inf	0.17	0.17	1
Human toxicity – HTP inf	2.97	0.35	1
Marine aquatic ecotoxicity – MAETP inf	0.27	0.25	1
Ozone layer depletion – ODP steady state	0.82	0.99	1
Photochemical oxidation – high No_x	0.66	2.29	1
Terrestrial ecotoxicity – TETP inf	0.28	0.18	1
Total	6.98	6.14	11
Average normalised impact	**0.63**	**0.56**	**1**

Sensitivity analysis

Process contribution, end-of-life modelling and data uncertainty were identified as important independent variables that could impact the dependent variables and thus overall LCIA under the system assumptions.

Data uncertainty and availability

Data uncertainty with a likely significant impact on results is the lack of LCI data for the wood preservation chemicals. A local timber treatment expert provided chemical composition and quantities of treatment required per cubic metre of timber, but impacts that could possibly occur when the treated product is disposed of were not accounted for. Similarly, no detailed LCI data were available for galvanised LGS. Global steel manufacturing processes in ecoinvent, including steel production, sheet rolling, zinc coating and metal working were combined and adjusted to approximate a local LGS product model. Metal working was included to represent the machining and press factory processes which produce profiled LGS truss components. This process contributes 36% to the LGS GWP and might be a slight overestimate as a result of the difference in general metal machining and LGS.

Although the Australian forestry models used reasonably represent local conditions, in order to better assess the impact of forestry on local land and water use, local LCI data would be required. In general, global LCI data are good enough for a general comparison, to assess trends and identify weak points in a system, but the calculated numbers should not be taken as absolute values. The work by Nebel et al.[34], on adapting European data for use in New Zealand, highlights the difficulty of using data from one country or region for another country that does not share common manufacturing resources. The latter can be especially difficult to assess in terms of appropriateness for an LCA practitioner.

End-of-life scenario discussion

Only one scenario was considered in this study: 100% material waste incineration. The assumption satisfies the reality of local wood waste treatment and scrap steel disposal. However, a study done by Blengini[35] showed that building material recycling has the potential to save between 18% and 35% on GWP over the building's life cycle.

Additional climate benefits of wood use can also be realised at the end of its life depending on biogenic carbon and GWP accounting approaches and by granting substitution benefits. In general, wood use can help reduce GHG effects by four main routes, which are closely interlinked: (1) carbon can be stored in forests and (2) wood products, (3) wood products can substitute for other products, thus using less fossil fuel during manufacturing, avoiding process emissions and fuel emissions through biofuel substitution, and (4) carbon dynamics in landfills.[7] Previous studies on the topic of wood substitution have found that the greatest potential for positively effecting climate change mitigation lies in increasing the amount of carbon stored in wood products and by substituting fossil fuels using wood energy or products that use a large amount of fossil fuel in their production.[28-30]

In this study, we chose a conservative approach to account for climate change benefits of wood use and substitution without accounting for carbon pools, carbon pool changes and substitution benefits to facilitate a relatively simple and easy direct comparison of the different roof truss systems and materials.

Conclusion

In both cradle-to-gate and cradle-to-grave analyses, the two timber alternatives – Biligom and South African pine truss systems – showed significantly lower environmental impact than LGS. For the smaller truss system, LGS had about twice the GWP impact of the timber systems and the normalised impact over all environmental indicators was about 40% higher. The benefit of biogenic carbon dioxide and low embodied energy present in timber proved to play a significant role in the GWP impact and could be further reduced if wood were used at its end-of-life to generate energy and substitute for fossil fuel use.

Overall, we have shown the potential advantage of using local timber products to reduce the environmental impact of the truss and building industry in South Africa. More local LCI data and research are required in order to promote and simplify direct system comparison in the local building industry and to better account for localised environmental emissions e.g. end-of-life fate of preservative treated timber. While better data would produce more reliable and robust absolute data, no changes to the general trends of this study are likely.

Acknowledgements

We thank the following persons for their insightful contributions and valuable data: Conrad van Zyl from Mitek South Africa, Spencer Drake, Biligom International, Doug Sayce, Lonza and Herman Aucamp. We gratefully acknowledge the Hans Merensky Foundation for providing study sponsorships which enabled the research to be undertaken.

Authors' contributions

P.L.C. was responsible for the article design, technical analyses and write-up; M.B. was responsible for the model analyses and write-up; C.B.W. supervised the research and revised the manuscript.

References

1. Wang L, Toppinen A, Juslin H. Use of wood in green building: A study of expert perspectives from the UK. J Clean Prod. 2014;65:350–361. http://doi.org/10.1016/j.jclepro.2013.08.023
2. US Energy Information Administration. How much energy is consumed in residential and commercial buildings in the United States? [homepage on the Internet]. No date [updated 2016 Apr 06; cited 2017 Apr 25]. Available from: http://www.eia.gov/tools/faqs/faq.cfm?id=86&t=1
3. Ding GKC. Sustainable construction – The role of environmental assessment tools. J Environ Manage. 2008;86(3):451–464. http://doi.org/10.1016/j.jenvman.2006.12.025

4. Petersen AK, Solberg B. Environmental and economic impacts of substitution between wood products and alternative materials: A review of micro-level analyses from Norway and Sweden. Forest Policy Econ. 2005;7(3):249–259. http://doi.org/10.1016/S1389-9341(03)00063-7

5. Werner F, Richter K. Wooden building products in comparative LCA: A literature review. Int J Life Cycle Assess. 2007;12(7):470–479. http://doi.org/10.1065/lca2007.04.317

6. Upton B, Miner R, Spinney M, Heath LS. The greenhouse gas and energy impacts of using wood instead of alternatives in residential construction in the United States. Biomass Bioenerg. 2008;32(1):1–10. http://doi.org/10.1016/j.biombioe.2007.07.001

7. Sathre R, O'Connor J. Meta-analysis of greenhouse gas displacement factors of wood product substitution. Environ Sci Policy. 2010;13(2):104–114. http://doi.org/10.1016/j.envsci.2009.12.005

8. Ximenes FA, Grant T. Quantifying the greenhouse benefits of the use of wood products in two popular house designs in Sydney, Australia. Int J Life Cycle Assess. 2012;18:891–908. http://doi.org/10.1007/s11367-012-0533-5

9. Oliver CD, Nassar NT, Lippke BR, McCarter JB. Carbon, fossil fuel, and biodiversity mitigation with wood and forests. J Sustainable For. 2014;33(3):248–275. http://doi.org/10.1080/10549811.2013.839386

10. Crickmay and Associates. South African lumber index: September 2016. Pietermaritzburg: Crickmay and Associates (Pty) Ltd; 2016.

11. Crafford PL, Wessels CB. The potential of young, green finger-jointed *Eucalyptus grandis* lumber for roof truss manufacturing. South Forests. 2016;78(1):61–71. http://dx.doi.org/10.2989/20702620.2015.1108618

12. Evans S. SA steel industry on the brink of collapse. Mail and Guardian [serial on the Internet]. 2015 August 29 [cited 2017 Apr 25]. Available from: http://mg.co.za/article/2015-08-27-sa-steel-industry-on-the-brink-of-collapse/

13. Spotlight on steel in South Africa. SA Roofing. 2015;70:14–15. Available from: https://issuu.com/trademaxpublications/docs/sar_june_2015_dload?e=7285681/13191678

14. Worth Z, Boyle C, McDowall WR. Combined life-cycle cost assessment of roof construction. Proc Inst Civil Eng Eng Sustain. 2007;160(4):189–198. http://dx.doi.org/10.1680/ensu.2007.160.4.189

15. Bolin CA, Smith ST. Life-cycle assessment of CCA-treated wood highway guard rail posts in the US with comparisons to galvanised steel guard rail posts. J Transp Technol. 2013;3:58–67. http://dx.doi.org/10.4236/jtts.2013.31007

16. United Nations. Millennium development goals and beyond [homepage on the Internet]. No date [updated 2015 Nov 10; cited 2017 Apr 25]. Available from: http://www.un.org/millenniumgoals/

17. Wilkinson B. South Africa – leading the green building charge in Africa [homepage on the Internet]. c2016 [cited 2017 Apr 25]. Available from: http://www.worldgbc.org/news-media/south-africa-leading-green-building-charge-africa

18. Curran MA. Life cycle assessment: Principles and practice [document on the Internet]. No date [updated 2006 May 01; cited 2017 Apr 25]. Available from: http://www.cs.ucsb.edu/~chong/290N-W10/EPAonLCA2006.pdf

19. Ciambrone DF. Environmental life cycle analysis. Boca Raton, FL: CRC Press; 1997.

20. ISO. Environmental management: Life cycle assessment: Principles and framework. ISO14040:2006.

21. ISO Environmental management: Life cycle assessment: Requirements and guidelines. ISO14044:2006.

22. Weidema BP, Bauer C, Hischier R, Mutel C, Nemecek T, Reinhard J, et al. Overview and methodology: Data quality guideline for the ecoinvent database version 3 [document on the Internet]. c2013 [cited 2016 Aug 16]. Available from: https://www.ecoinvent.org/files/dataqualityguideline_ecoinvent_3_20130506.pdf

23. ArcelorMittal. ArcelorMittal South Africa [homepage on the Internet]. No date [cited 2017 Apr 25]. Available from: http://southafrica.arcelormittal.com/

24. South African Wood Preservers Association. CCA treated timber [document on the Internet]. No date [cited 2017 Apr 25]. Available from: http://www.sawpa.co.za/documents/THE%20SAFETY%20OF%20CCA%20TREATED%20TIMBER%20Revision%20final.pdf

25. Niyobuhungira RV. An investigation of CCA-treated wood in informal caterers' fuel stocks and related airborne arsenic in the Cape Town region [thesis]. Cape Town: University of Cape Town; 2012.

26. South African Waste Information Centre. Scrap metal fact sheet [document on the Internet]. No date [cited 2016 May 16]. Available from: http://sawic.environment.gov.za/documents/5329.pdf

27. openLCA. GreenDelta [homepage on the Internet]. No date [cited 2017 Apr 25]. Available from: http://www.openlca.org/

28. Perez-Garcia J, Lippke B, Comnick J, Manriquez C. An assessement of carbon pools, storage, and wood products market substitution using life-cycle analysis results. Wood Fiber Sci. 2005;37(Corrim special issue):140–148.

29. Gustavsson L, Pingoud K, Sathre R. Carbon dioxide balance of wood substitution: Comparing concrete- and wood-framed buildings. Mitig Adapt Strategies Glob Chang. 2006;11(3):667–691. http://dx.doi.org/10.1007/s11027-006-7207-1

30. Lippke B, Jonhson L, Wilson J, Puettmann M. Life-cycle environmental performance of renewable building materials in the context of residential construction [document on the Internet]. c2011 [cited 2017 Apr 25]. Available from: http://www.sefs.washington.edu/research.corrim/pubs/reports/2010/phase2/Ph2_Main_Report.pdf

31. Cherubini F, Peters GP, Berntsen T, Stromman AH, Hertwich E. CO_2 emissions from biomass combustion for bioenergy: Atmospheric decay and contribution to global warming. Glob Change Biol Bioenergy. 2011;3(5):413–426. http://dx.doi.org/10.1111/j.1757-1707.2011.01102.x

32. Helin T, Sokka L, Soimakallio S, Pingoud K, Pajula T. Approaches for inclusion of forest carbon cycle in life-cycle assessment – a review. Glob Change Biol Bioenergy. 2013;5:475–486. http://dx.doi.org/10.1111/gcbb.12016

33. US Environmental Protection Agency. Framework for assessing biogenic CO_2 emissions from stationary sources [document on the Internet]. c2014 [cited 2017 Apr 25]. Available from: https://yosemite.epa.gov/sab/sabproduct.nsf/0/3235DAC747C16FE985257DA90053F252/$File/Framework-for-Assessing-Biogenic-CO2-Emissions+(Nov+2014).pdf

34. Nebel B, Alcorn A, Wittstock B. Life-cycle assessment: Adopting and adapting oversear LCA data and methodologies for building materials in New Zealand [document on the Internet]. c2011 [cited 2017 Apr 25]. Available from: http://maxa.maf.govt.nz/forestry/publications/lca-materials.pdf

35. Blengini GA. Life-cycle of buildings, demolition and recycling potential: A case study in Turin, Italy. Build Environ. 2009;44:319–330. http://dx.doi.org/10.1016/j.buildenv.2008.03.007

Perspectives of wild medicine harvesters from Cape Town, South Africa

AUTHORS:
Leif Petersen[1,2]
Andrew M. Reid[1]
Eugene J. Moll[3]
Marc T. Hockings[4]

AFFILIATIONS:
[1]Sustainable Livelihoods Foundation, Cape Town, South Africa
[2]National Research Foundation Centre of Excellence in Food Security; Poverty, Land and Agrarian Studies, University of the Western Cape, Cape Town, South Africa
[3]Biodiversity and Conservation Biology, University of the Western Cape, Cape Town, South Africa
[4]School of Earth and Environmental Sciences, University of Queensland, St Lucia, Brisbane, Australia

CORRESPONDENCE TO:
Leif Petersen

EMAIL:
leif.petersen@livelihoods.org.za

DATES:
Received: 22 Sep. 2016
Revised: 12 Jan. 2017
Accepted: 22 May 2017

KEYWORDS:
conservation; wild harvest; sustainability; conflict, medicinal plants

HOW TO CITE:
Petersen L, Reid AM, Moll EJ, Hockings MT. Perspectives of wild medicine harvesters from Cape Town, South Africa. S Afr J Sci. 2017;113(9/10), Art. #2016-0260, 8 pages. http://dx.doi.org/10.17159/sajs.2017/20160260

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
None

Cape Town is a fast-growing cityscape in the Cape Floristic Region in South Africa with 24 formally protected conservation areas including the World Heritage Table Mountain National Park. These sites have been protected and managed as critical sites for local biodiversity, representing potentially one-third of all Cape Floristic Region flora species and 18% of South Africa's plant diversity. Cape Town is also inhabited by a rapidly growing culturally and economically diverse citizenry with distinct and potentially conflicting perspectives on access to, and management of, local natural resources. In a qualitative study of 58 locally resident traditional healers of distinct cultural groups, we examined motivations underlying the generally illicit activity of harvesting of wild resources from Cape Town protected areas. Resource harvester motivations primarily link to local economic survival, health care and cultural links to particular resources and practices, 'access for all' outlooks, and wholesale profit-seeking perspectives. We describe these motivations, contrast them with the current formal, legal and institutional perspectives for biodiversity protection in the city, and propose managerial interventions that may improve sustainability of ongoing harvest activities.

Significance:

- The study reveals, for the first time in the Cape Floristic Region, informal economy viewpoints on terrestrial nature and how its direct use has important economic and cultural roles – specifically in wild medicine harvesting and trade.
- We contrast the formal and informal approaches to nature conservation in the city and propose new considerations for conservation managers.

Introduction

Wild-harvested medicines form part of the historical and contemporary fabric of South African society, and are used by at least 27 million consumers in a largely complementary manner to Western medicine.[1] The industry represents a 'multimillion rand hidden economy'[2] previously estimated to be worth ZAR2.9 billion per year nationally.[1] The bulk of materials traded as traditional medicines are wild harvested from natural populations in wild habitats, in some cases resulting in adverse ecological impacts through species decline.[1,3]

Since colonial settlement, and now sympatric to this long-established traditional industry, an organised process of legislating and developing policy for conservation and protected area management has been implemented in South Africa, overseen by conservation agencies in a now democratically elected government. These activities have led to the establishment of a legislatively protected and complex array of public and privately managed biodiversity areas, which form the mainstay of modern national and international conservation strategies.[4] In part due to historical circumstance, ownership or custodianship and the varied management requirements for its diverse landscapes, the formal conservation framework in South Africa reflects a range of concurrent competences, with national and provincial responsibilities, and regionally differing policies with regard to species management. Within this context of formally structured conservation systems, Shackleton[5] observes that much of the ongoing wild harvest of biological resources from wild habitats takes place in a management vacuum – exacerbated by the lack of a clear government department specifically responsible for sustainable resource use, commonly poor institutional capacity in existing departments, and erosion in traditional authority leadership and communal tenure[6] which might otherwise play a role in overseeing such activities. The lack of comprehensive oversight of this culturally important harvest has increased conflict between the formal custodians of biodiversity such as protected area managers and many direct users of wildlife resources. This conflict is increasing in Cape Town – the urban centrepiece of the Cape Floristic Region – in which 448 biological species are harvested from public and wild areas in and around the city.[7] Wild resource harvesting presents new challenges for city conservation officials and is a stated local conservation issue.[8]

Whilst protected area management goals and strategy are clearly articulated and practised in South Africa, less is known about the perspectives of those operating in the informal natural resource trade. Using Cape Town as a study site, we aimed to better understand wild resource harvester motivations for their extraction activities. Conducting research to better understand local harvester livelihood and cultural outlooks towards formal conservation approaches may better inform conservation strategy and natural resource use approaches in the city.

Background

Protected areas and natural resource harvesting

Contemporary South African conservation management is administered through a variety of agencies in the three layers of government – national, provincial and local, with around 6% of South Africa's land surface under some form of legal conservation protection[9] in nearly 500 state-operated protected areas[10]. South Africa has a further well-developed private nature reserve network based on tourism, game farming and ranching.[11] A comprehensive

legislative and policy environment linked to the conservation and management of natural resources has been established and implemented (summarised in Appendix 1 of the supplementary material).

There is increasing debate around the role and format of protected area management systems, and the place of people and resource use related to them.[12] In Africa and South Africa, where poverty is commonplace, these debates are increasingly acute. Despite South African state efforts to protect and manage conservation landscapes, the legalistic and management frameworks put into place have not necessarily led to a decline in local community reliance on natural resources harvesting both within and outside of protected areas.[13] In some respects, consumptive resource reliance is increasingly pronounced in financially poor rural areas and includes local vulnerability reduction strategies such as collecting fuelwood for heating and cooking, with estimates of reliance on fuelwood as a primary energy source as high as 92% in Bushbuckridge (a town in the province of Mpumalanga)[14], and up to 76% in rural Eastern Cape[15]. Harvesting of wild 'edibles' including marula fruits[16], mopane worms and bushmeat[17], is also commonplace. Furthermore, woodcutters and carvers remain largely reliant on wild harvested materials.[18]

A large but almost entirely informal industry is the trade in wild harvested medicine which directly employs at least 133 000 people,[1] including culturally important traditional healers.

Traditional healers

The varied cultural practice of traditional healing is generally a holistic approach to patient well-being, and considers both spiritual and physical welfare.[19] Traditional healers vary in practice, and include groups such as predominately black South African inspired *amaxwhele* who work through a medium of dreams and visions to diagnose patient life-challenges and misfortune, through to Rastafarian-styled bush doctors who provide specialist knowledge services and treatments including those for symptomatic illnesses that could be considered familiar to Western practice.[20] Predominant in the Western Cape, Eastern Cape and Northern Cape Provinces, many Rasta herbalists claim descendancy links with indigenous Khoi citizens and draw on a diverse pharmacopoeia – including medicines of Cape Dutch, Zulu, Xhosa and European provenance – to treat physical ailments.[20] The integration of Rasta philosophy and medicinal plant knowledge, embodied in Rasta herbalist practices, emphasises self-sufficiency, attunement with nature and holistic healing.[20]

Related to such practices lie traditional medicine services for treating cultural afflictions.[21] In this case, illnesses commonly termed *Idziso* are considered to consume their victims through social misfortune, illness and death.[22] Removing *Idziso* (commonly known as 'African poison') requires treatment by powerful professional healer-priests such as *amagqirha* who use specialised traditional herbal medicines. Such specialists operate at the interface of practising religion, magic and medicine,[22] maintaining a supernatural contact with ancestral spirits who in turn advise the healer on the causes of patient misfortunes and prescribe the required treatments.

South Africa's widespread traditional healing services support a wild medicine industry that effectively operates in addition to, and sometimes in preference to, Western medical systems.[23] Traditional healing practices are foundationally different from Western medicine – virtually all healers are trained through oral apprenticeships by elder healers; medicinal ingredients are commonly wild-harvested by healers or traders and transported by foot and informal minibus taxi; and medicines are traded in cash from street or home-based treatment rooms.

Wild medicine industry

South Africa's wild medicine industry is vibrant, widespread and growing, and utilises an estimated 20 000 tonnes of biological materials per year (primarily from the provinces of KwaZulu-Natal, Limpopo, Mpumalanga and the Eastern Cape). South Africa's rich literature includes Mander et al.[1,24] describing the KwaZulu-Natal and national trades; Botha et al.[3] examining the Lowveld savannas; Dold and Cocks[19] investigating the Eastern Cape harvest and trade; and Williams et al.[25] exploring the medicinal trade in Gauteng. In response to the growing academic and scientific recognition of these historical linkages between communities and natural resources, there has been a gradual shift in policy and legislative approaches to recognise these activities. This recognition has meant that legislation, policy and institutional efforts have increasingly been framed to allow for potential collaborative opportunities in local-level natural resources management. Legislation such as the *National Environmental Management Act*[26] specifically recognises and articulates the need for

> *equitable access to environmental resources, benefits and services to meet basic human needs and ensure human wellbeing in accordance with diverse interests, needs and values of all interested and affected parties, including recognition of all forms of traditional and ordinary knowledge.*[26]

For example, in response to a land claim in the Kruger National Park from the Makuleke community[27], South African National Parks have attempted a specific local collaborative resource management strategy. Such approaches increasingly dominated the conservation debate around the time of South Africa's transition to democracy[12] and have subsequently broadened state considerations of public rights and access to resources.

Whilst the legislative and policy environment around the formal protection and management of biodiversity moves away from colonial strategy and increasingly reflects broader community ideals, some authors have highlighted limitations to these approaches. Firstly, as pointed out by Cousins[28], there is a concern that much legislation and titling in South Africa reflects Western notions of land ownership that are largely unrepresentative of African tenure systems. Secondly, many community-based natural resource management ventures are based on formal property rights which in turn consolidate or marginalise different community groups.[29] Thirdly, as highlighted by Shackleton[5], governmental management responsibility for mainstreaming community access to natural resources is largely fragmented and weakly implemented.

Considering the reliance on wild-harvested natural resources for health, well-being and economic purposes in South Africa, and the challenges faced for the nation's broader economic development, there is potential for conflict between biodiversity protection and utilisation perspectives. Certainly, if wild resource harvesting from natural habitats takes place unchecked, the activity could have negative impacts on South Africa's biodiversity, the harvester livelihoods and public health (for those who rely on wild-harvested medicines). By interviewing 58 wild resource harvesters in Cape Town, we aimed to develop an understanding of their perspectives towards natural resources, protected areas and their stated motivations for conducting harvests.

The research site: City of Cape Town

Cape Town is the capital of the Western Cape Province of South Africa. The well-established historical city occupies land between central Table Mountain National Park and beaches fronting the South Atlantic Ocean. Moving eastwards, the city sprawls onto a large sandy plain of 232 working class 'townships' and informal settlements[30] locally known as the Cape Flats. The city accommodates 3.84 million people[31] but is a fast-growing centre, largely as a result of the arrival of up to 13 000 predominantly economic migrants per month from the Eastern Cape[32]. Formal unemployment exceeds 60% in some parts of the city.[33]

Existing research within Cape Town has documented specific aspects of informal trade in local wild-harvested resources. For example, Rebelo[34] noted Proteaceae species being wild-harvested at low levels on the Cape Peninsula; Cowling and Richardson[35] recorded sour figs (*Carpobrotus* spp.), honeybush tea (*Cyclopia* spp.) and buchu (*Agathosma* spp.) as prominent local foods; Turpie et al.[36] noted harvesting and informal trade of firewood (introduced *Acacia* spp.); Clark et al.[37] highlighted the importance of subsistence fisheries to local economies; and an area of considerable government focus has been local illicit harvesting and export of abalone (*Haliotis* spp.) linked to criminal syndicates[37,38]. Furthermore, there exists a large but poorly documented

cash industry of wild medicines operated by diverse cultural groups reliant on wild-harvested biological materials. At least half of all locally harvested species recorded by Petersen et al.[7] were utilised as wild medicines; Loundou[39] noted the retailing of 170 medicinal plant species within the city and Nzue[40] documented the local harvest of 52 species as medicines. Because of conservation legislation and land tenure, many medicinal plant harvests were illegal. As such, this culturally important and prominent informal economy business activity was investigated.

Methods

Research was undertaken in four Cape Town working-class residential communities in which informal business activity predominates. These communities were selected to be representative of the city's economically marginalised population groups, levels of unemployment, population ethnicity (black and coloured South Africans), urban characteristics (from slum settlements to formalised housing) and geographical spread (Table 1).

Between July and November 2011, we traversed the selected suburbs on bicycle and foot, conducting an informal enterprise census by enumerating natural resource businesses street by street within the suburb boundaries. Some business activities were visually obvious, such as those with signage, whilst others were in unmarked households and identified through participant referrals. From this cohort, traditional healers were invited to participate in a 40-min interview within their business. All healers who readily admitted to harvesting their own wild medicines were interviewed using qualitative open-ended questions on perceptions of formal conservation strategy, individual motivations for wild resource harvesting, challenges for harvesting activities and the business of traditional healing. No prompts were given, in order to limit any researcher bias or influence in answering. To ensure accurate and reflective responses, a culturally representative and multilingual investigatory team was trained by the lead researcher (L.P.) and participated in interviews. Each participant was informed of the objectives of the research and their consent secured. Trust was gained through the extensive mobilisation process whereby the researchers became well known in the local community and word-of-mouth connections were made between traditional healers as to the intentions of the research team.

All responses were anonymised and documented into Microsoft Excel tables for comparative review. As individual interviews were transcribed, responses were grouped into broad themes of harvester perspectives, which were subsequently discussed and confirmed in a focus group of eight traditional medicine practitioners in early 2012.

Ethical clearance for the study was obtained from the University of Queensland.

Results

A total of 58 wild resource harvesters were interviewed from three predominant traditional healer groupings. None had regular formal sector employment and all highlighted regular (monthly to quarterly) local and regional illicit resource-harvesting activities. All respondents processed and retailed their products through regular or occasional cash sales via personal networks or street vending within local markets (Table 2).

The participants demonstrated a variety of resource access and utilisation viewpoints pertaining to local biodiversity, justifying their harvesting, trading and consumptive activities using a variety of explicit claims and perceived entitlements. Analysis of the field interviews allowed for basic categorisation into a range of utilisation motivations as explicitly claimed by the participants, with many presenting various justifications for their activities (Table 3).

Predominant perspectives

In the following section, both predominant economic (survivalists and profit seekers) and social themes (cultural requirements, access for all and indigenousness) are similarly grouped together.

Group 1: 'I need this money as I don't have a job' – Economic survival

Over 90% of respondents stated that they conducted wild resource harvesting primarily as an economic survival strategy by generating cash income from sales of wild-harvested resources for livelihood support. Analysis of average household sizes and income streams revealed that natural resource sales were indeed an important livelihood activity for respondents (Table 4).

Table 1: Urban localities in Cape Town sampled for wild harvested natural resource based businesses and traditional healers

Name	Urban characteristic	Population†
Capricorn	Formalised, electrified urban settlement comprising patron-funded three-room brick houses	18 270
Overcome Heights	Largely unserviced informal settlement of owner-built shack dwellings	11 587
Seawinds	Formalised, electrified urban settlement comprising privately built and patron-funded brick houses	7689
Sweet Home	Largely unserviced informal settlement of owner-built shack dwellings	16 000
Total population sample		**74 977**

†Population data derived from Statistics South Africa[41] updated by Census Plus[42] and high-resolution aerial photo dwelling counts.

Table 2: Typologies of traditional healers interviewed to determine motivations for wild harvesting and trading in natural resources

User	Method of trade	Description	Number interviewed
Amagqirha	Private consultation with clients	Traditional healers / shamans who stand between ancestral spirits and patients and use spiritual abilities for healing	36
Rastafarian herbalists	Street trade to the public	Pan-Africanists with proclaimed religious links to Ethiopia and informally acquired specialised ethno-botanical healing knowledge	13
Amaxwhele	Private consultation with clients	Local residents with informally acquired specialised ethno-botanical healing knowledge	9

Table 3: Explicit natural resource utilisation claims of traditional healer research participants

Utilisation perspective	Description	Explicit claim	Percentage of respondents
Economic survival	Harvesting to bolster household resource and income security	53	92%
Cultural requirements	Harvesting to fulfil well-established and continuing demands linked to family or community history	52	89%
Access for all	A perspective of generalised open-access rights to landscapes to collect resources	46	80%
Indigenousness	Individuals claiming a historical continuity with pre-colonial societies that developed on their territories; consider themselves distinct from the state	13	22%
Profit seeking	Harvesters who seek to generate increasing personal wealth from natural resource products through illicit harvesting and informal trading opportunities	5	8%

Table 4: Economic contribution of wild harvested resources to traditional healer households

User type	Average household size	Total household income per month, in ZAR (USD)	Resource-based income per month, in ZAR (USD)	Percentage contribution of wild resources
Amagqirha	4.8	2916 (243)	2364 (197)	81%
Rastafarian herbalists	3.1	6708 (559)	4008 (334)	59%
Amaxwhele	3.0	2820 (235)	1380 (115)	49%

An average exchange rate for 2012 of USD1 = ZAR12.00 was used.

In all cases, despite the potential illegality of the activity, half or more of household income was based on wild-harvested resources. For the *amagqirha* in particular, the economic reliance on wild resources was considerable, especially in light of the large average household size.

Many respondents claimed to harvest and trade wild-harvested resources because of a lack of alternative income-earning opportunities, retrenchment, and the need for cash income. Few were exclusively reliant on this income stream – more than half (55%) of the respondents came from households which received social grants from government (either old-age pensions or child-support grants), whilst 43% lived in households with income from non-natural resource related work. Considering the high relative contribution of natural resource incomes in all households, it is evident that the wild harvest of natural resources enhances economic survival for these residents, certainly propelling many above minimum living levels of ZAR744 (USD62) per person per month.[43] Additionally for some Rastafarian healers, there was considerable personal consumption of harvested wild resources as medicine – effectively subsidising their household health care and allowing them to save financial resources from other economic activities. For those living on the fringes of the cash economy, the reliance on wild-harvested resources was substantial in terms of both income opportunities and personal consumption, and these products served as an important safety net.[44,45]

Group 2: 'I'm in business to make money' – Profit seeking

Building on economic survivalist motives, the relatively rapid emergence of South Africa's modern economy in the last few hundred years, coupled with cash demand for culturally important items, has created large commercial markets for wild-harvested resources.[1] In this case, capitalistic entrepreneurs – commonly utilising traditional knowledge – seek to derive increasing commercial benefit from wild harvesting of natural resources. Although similar to motives of economic survival, two entrepreneurial respondents were utilising motor vehicles, plant shredders and other dedicated machinery to harvest and process wild resources, and were selling on bulk quantities of processed resources to other healers. Like local extraction of *Haliotis midae*[38], these activities were increasingly clandestine and organised compared to the activities of other harvesters. Potential profiteering activities identified by this research include the extraction of the indigenous medicinal plant *Tulbaghia capensis* (wild garlic) of which 16 000 bulbs were confiscated from up to 50 known (and repeatedly arrested) individuals illicitly harvesting in the city-managed 300-ha Tygerberg Nature Reserve over 2 years from 2010 (Glanville P 2011, personal communication, November). Law enforcement records in the reserve show that some arrests linked to this activity realised hundreds of bulbs confiscated from individual collectors at any one time – greatly exceeding quantities for personal or cultural use typified by the resource stockpiles of the majority of respondents, with produce reportedly heading to local informal trading markets throughout and beyond the city.

Group 3: 'We need the herbs to heal the people' – Health and well-being

For many South Africans, plant medicines are sought as stabilisers and proactive responses to the precarities and uncertainties of everyday life: the need to secure employment, attract a potential partner, or realign one's conduct in relation to past generations of family.[46,47] These volatilities are often heightened in densely populated urban areas where people may be exposed to physical and psychological trauma or ill-health as a result of disease (TB/HIV being examples), substance abuse, or diets founded on nutrient-poor foods high in sugar and fat. In addition, aspirations to success and heightened prestige in highly competitive urban contexts are threatened by the limited availability of, and access to, necessary resources and opportunity. Contemporary urban citizens with familial histories of medicinal plant use continue to draw on these practices holistically to help promote their well-being and greater future prosperity, by cleansing themselves, family members and their immediate surroundings.[21,48,49] Traditional healers serve as a vital source of medicinal plants and knowledge for health and well-being. Nearly 90% of respondents highlighted the cultural importance of conducting the wild harvesting of medicinal resources themselves. This cultural influence remains deeply rooted, with the *amagqirha* in particular describing themselves as 'their ancestors' servants', suggesting that a set of higher forces controlled their day-to-day traditional practices, necessitating wild resource harvest practices and use.

In addition to their use as well-being enhancing medicines, plants are commonly sought out and relied on to treat physical ailments such as diarrhoea, arthritis and high blood pressure.[1,47] Importantly, the introduction and rise of commercial pharmaceuticals in South Africa

has not radically altered traditional healer beliefs, rather these newer technologies and tools are commonly used in a complementary, alternative and integrative manner with traditional medicine.[48] Similar trends involving the fluid integration of multiple treatment approaches are seen around the world.[50-52] One traditional healer respondent illustrated this point stating that 'the clinic is for Western problems' with another stating that 'there are no tablets for a job at the clinic'. The health and well-being requirement fosters continued consumer demand for locally occurring wild resources and remains an important driver for local resource harvesters of all backgrounds.

Group 4: 'These herbs belong to the people' – Access for all viewpoints

As much as 80% of respondents highlighted their perceived rights to access local protected areas and natural resources for the purposes of consumptive harvesting. The basis of this viewpoint is a perspective of open access rights to resources, with wild habitats considered as important sources for biological materials. Many of Cape Town's black African citizens (including 44 of the 45 black South African *amagqirha* and *amaxwhele* in this research) were born in apartheid-established Bantustans in the current-day Eastern Cape Province. Historically, these territories were managed under state sanctioned regimes of local traditional leadership, and in a complex mix of tenure systems including trusts, quitrent, freehold, communal and tenancy arrangements, with little freehold land.[53] Before and after South Africa's democratic transition, these localities have remained largely under local community control, with many openly used as sources for wild medicinal plants under systems of community control. For many recent Cape Town migrants from these localities, utilisation outlooks towards land and natural resources are common, with local state property and protected areas reportedly being seen as new harvest sites for collections. However, in this case, fencing, conservation management and law enforcement have become the 'occupational hazards'[5] in a landscape under a different form of (state) control.

Group 5: 'We are the original people and this land is for all of us' – Indigenousness

The United Nations[54] generally considers indigenous peoples to be those who, having a historical continuity with pre-colonial societies that developed on their territories, consider themselves distinct from other sectors of the societies now prevailing on those territories. Within Cape Town, claims of indigenousness primarily emanate from those claiming descendancy from the Khoi peoples present at the time of European settlement but largely eliminated from the City as a result of various colonial practices and cultural suppression.[55]

The seven interviewed Rastafarians of coloured ethnicity highlighted immediate local indigenousness and concurrent herbal knowledge as an important political justification for their activities. Rastafari herbalists participating in the study were keenly aware of the historical legacy of colonial oppression and claims of indigenousness were expressed and justified through proclaimed and evidenced knowledge of Khoi culture, heritage and cultural ascription.[56] For those making indigenous claims, mainstream societal laws, land ownership and modern conservation protection methods were considered non-applicable and politically unjustifiable. The emotional significance of these perceived entitlements with respect to wild gathering plants from the local Table Mountain National Park was embodied in a statement by one Rastafarian participant: 'I need these herbs for my heart to be secured'. Another participant framed access to medicinal plants in terms of post-apartheid reparation: 'These herbs are our inheritance. District Six was returned to the people who were moved from there and these herbs must be returned to us.'

Discussion

Formal protected areas and informal use

South Africa's formal system of protected areas primarily serves the broader interests of society and the economy through the protection of biodiversity, scenery and watersheds, and the promotion of tourism and other attributes.[9] Despite legal protection, Cape Town conservation landscapes are under increasing threat from high rates of local population growth and urbanisation, alien plant invasion and climate change.[57] Adding pressure is a growing local informal economy based on wild-harvested natural resources.[7] Individuals conducting these harvests demonstrated cultural and livelihood motivations that differ from the perspectives and motivations of the formal framework of conservation planning and management, expressed largely through cultural and economically driven consumptive use. An overarching difference underpinning various harvester motivations relates to perceived 'usage rights' over resources growing on state and private land, which for many in the study are viewed as belonging to supreme deities.[58]

Importantly, despite individual harvest impacts (which in some cases were considerable) all research participants demonstrated considerable appreciation of the broader merits of biodiversity conservation. However, additional to this appreciation was a range of perceived user rights to resources underpinned by various economic, cultural and historical experiences and expressed in the widespread wild harvest and trade of hundreds of species of flora and fauna from within the City.[7] These largely informal 'livelihoods or culturally driven' activities potentially conflict with dominating 'protection driven' conservation strategy. Figure 1 presents a theoretical framework representing how participants in Cape Town's informal traditional medicine economy, underpinned by various economic and anthropogenic drivers, interpret local nature in contrast with formal approaches.

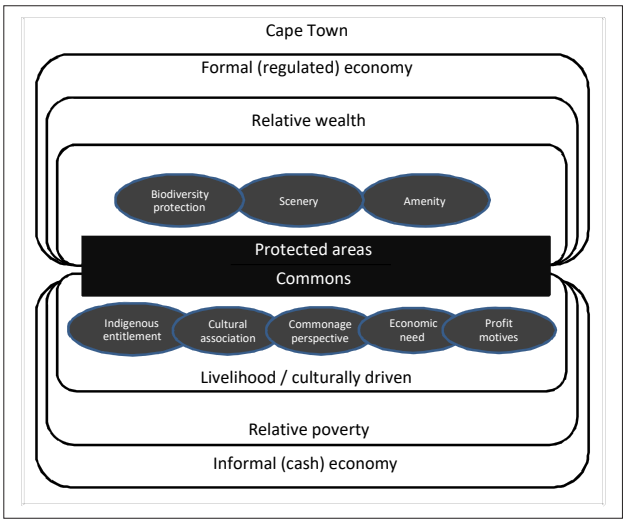


Figure 1: Theoretical framework highlighting anthropogenic perspective differences for Cape Town nature.

Originating in the informal economy under conditions of poverty, the harvesters of wild resources conduct this activity to satisfy health and well-being demands in local cash markets. The medicinal plant harvest and trade makes an important contribution to economic livelihoods, without which 76% of research participants would be living in conditions of financial poverty. Many interviewed in this research considered the formal, legalistic and protection driven outlook towards nature conservation as a middle-class interpretation of how nature should be maintained, which was considered culturally insensitive and ignorant of their lived reality of social and economic marginalisation.

Under present conditions, the intersection of formal conservation practice with natural resource extraction for the growing informal economy will inevitably lead to conflict. The formal local conservation approach to ecological systems through legislated protected areas, 'fines and fences', and primarily non-consumptive use is challenged by the views and activities of traditional healer participants in this research.

Evolving health-care practices and natural resource demand

The highly entrenched and growing demands for wild harvested medicines in South Africa[1-3] and the largely complementary role to Western medicine[1] means consumer preferences for this service and

related products are unlikely to decline in the near future. Further, the perspectives of various groups of natural resource harvesters and traders require acknowledgement by state conservation and natural resource agencies. In adapting to such use, programmes of community-based natural resource management have been implemented under broader ecological and economic sustainability objectives such as those outlined in the Secretariat of the Convention on Biological Diversity.[59] However, post-colonial indigenous-linked cultures and practices combined with the growing influence of cash, trade and technology makes for a very different relationship between 'traditional' people and nature when compared with historical situations in less densely populated contexts. Contemporary relationships involving indigenously rooted practices, practitioners and natural resources bring perceptions of ecologically friendly local communities, as commonly portrayed in popular media, into question.[60] As highlighted by Cocks[61], in many respects, the worldviews, cultural values and knowledge of large sectors of the population in [South] Africa can no longer be classified as 'traditional' nor as representative of Western culture. Rather, the shifting arrangements of life in contemporary South Africa – resulting in part from accelerated urban migration, unemployment and disease – stimulate dynamic and adaptive responses from individuals in order to secure livelihoods, in this case, through the harvest and trade of natural resources.

It was apparent in the local setting that traditional healer outlooks towards biodiversity appear largely economically and culturally entwined. This blending of motivations makes for complex and subjective arguments both for and against local consumptive wild resource utilisation. Figure 2 demonstrates how the influence of the cash economy has stimulated an evolution of wild resource harvesting and trade from more 'traditional' drivers within the context of cultural, subsistence or indigenous claims towards a spectrum of increasingly informal livelihood and income-driven informal and entrepreneurial motivations. The influence of cash motives into these harvest activities can evolve this 'traditional' harvesting into an 'informal' economy activity driven by cash and profit, and which include increasingly entrepreneurial collectors.

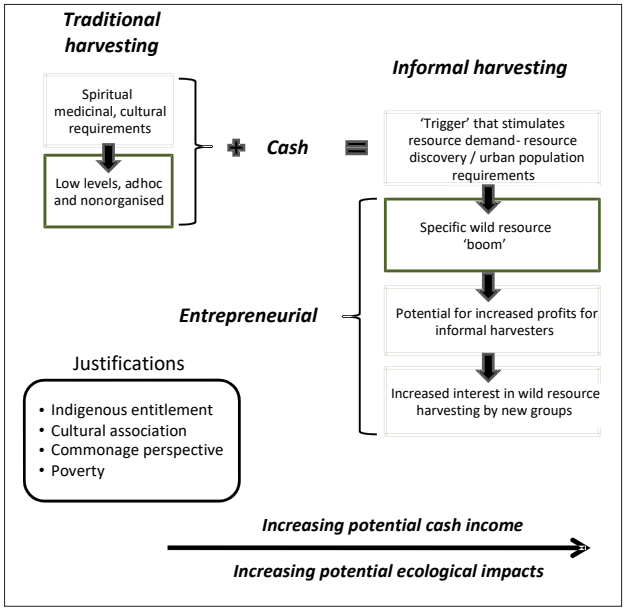


Figure 2: The evolution of illicit natural resource harvesting motivations in Cape Town residents.

In Cape Town, the drivers of natural resource harvesting for informal economy trade are evolving as the influence of cash markets in the sector grows. Coupled with this increase are greater potential impacts on conservation and protected area management. Thus, despite Cape Town's well-established, formalised, legally enforceable and ecologically critical protected area network, in light of anthropogenic drivers of poverty and cultural demands there remains a strong likelihood that wild resource harvesting from these areas will continue and may grow.

The conservation challenge

The variety of resource utilisation perspectives demonstrated by traditional healers, and evolving drivers of harvest, present a growing challenge for state attempts to balance biodiversity protection with local economic development needs. That said, the local situation reflected in this research is not necessarily unique – it reflects the deep political and social complexity inherent in conservation work in developing countries.[60]

In the absence of any one specific South African government agency specifically mandated to deal with natural resource use (as raised by Shackleton[5]) there is a need for relevant conservation agencies to consider the drivers and impacts of this activity. Within Cape Town, this consideration includes engagement with these issues by local authorities (Cape Town Nature Conservation), provincial (CapeNature) and national (South African National Parks) bodies – all of which directly manage biodiversity resources in and around the city. Practically, from their formal protected area management perspective, more assessments of local ecology are needed at smaller (reserve or species) scales[62] to better understand and monitor harvest impacts and to inform management responses to harvest risks. Where biodiversity risks are pronounced from wild resource collections, stricter protected area boundary maintenance coupled with basic social relief services as suggested by Wilshusen et al.[60] could be considered – such as raising local awareness of existing social protection grants and measures. In circumstances of genuine and provable local cultural ties to resources, the negotiation of legitimate and binding agreements for controlled harvesting could be undertaken on legally appropriate landholdings – although the number and diversity of traditional healer types and resource claimants will be highly problematic to manage or enforce with limited state resources.

Beyond conservation managers there is a need for municipal parks, gardens and public space managers to engage in practical land management activities such as encouraging publicly accessible harvest projects and landscapes within the urban landscape. As seen with the rise in cash trade, patterns of reliance on, and trade in, wild harvested medicines are not impervious to adaptation and change (see also Botha et al.[3], Dold and Cocks[63]). Whilst wild harvesting remains the current healer preference, it is possible that alternative strategies for resource collection, such as cultivation schemes, may gain local support. Although traditional healer groups are ethnically diverse and varyingly motivated, conservation agency supported planting of culturally important species could take place where suitable landholdings can be identified. Allowing harvests of these biological materials under an 'open-access' regime (as per current largely illicit collections) may allow divergent harvester groups the opportunity to continue their trade without necessarily compromising the biological integrity of local protected areas.

Conclusions

Contemporary protected areas in Cape Town conserve important biodiversity for South Africa. However, beyond the conventional scientific rationale for their protection, they are also seen from perspectives of diverse origins, with some considering consumptive use such as wild resource harvesting from these areas as an economic and cultural necessity. These different ways of understanding and appreciating nature underpin an emerging resource use conflict within the City. The geographical constraints of Cape Town combined with the increasing fragility of local biodiversity and emerging cash drivers of wild resource harvesting mean that, from a biodiversity protection perspective, unconstrained consumptive usage of wild resources is ecologically problematic. Importantly, enhancing understanding of harvester and alternative views of conservation beyond this study could assist in devising more inclusive and reflective conservation management practices – even for non-conservation lands that could be utilised for biodiversity business. Through such activities, the development of alternative livelihood opportunities around the culturally and economically important harvest of natural resources is required in ways that will not compromise ecological and industry sustainability in a dynamic and changing Cape Town society.

Acknowledgements

We thank Andrew Charman, Ray Collins and colleagues at the Sustainable Livelihoods Foundation.

Authors' contributions

L.P. was responsible for the literature review, methodology, field work, analysis and write-up; A.M.R. was responsible for field work and analysis; E.J.M. was responsible for the literature review, methodology and revision; M.T.J. was responsible for the literature review and supervision.

References

1. Mander M, Ntuli L, Diedrichs N, Mavundla K. Economics of the traditional medicine trade in South Africa. FutureWorks Report for Ezemvelo KZN Wildlife; 2007. Unpublished report.
2. Cunningham AB. Herbal medicine trade: A hidden economy. Indicator S Afr. 1989;6:51–54.
3. Botha J, Witkowski ETF, Shackleton C. Market profiles of medicinal plants in the Lowveld, South Africa. Environ Conserv. 2004;31:38–46. https://doi.org/10.1017/S0376892904001067
4. Hannah L, Midgley G, Andelman S, Araújo M, Hughes G, Martinez-Meyer E, et al. Protected area needs in a changing climate. Front Ecol Environ. 2007;5(3):131–138. https://doi.org/10.1890/1540-9295(2007)5[131:PANIAC]2.0.CO;2
5. Shackleton C. Will the real custodian of natural resource management please stand up. S Afr J Sci. 2009;105(3–4):91–93.
6. Cousins B, Hall R. The potential and limits of rights-based approaches to securing land tenure in rural South Africa. In: Langford M, Dugard J, Madlingosi T, Cousins B, editors. Symbols or substance? The role and impact of socio-economic rights strategies for Africa. Cambridge: Cambridge University Press; 2015.
7. Petersen LM, Moll EJ, Collins RJ, Hockings MT. Development of a compendium of local, wild-harvested species used in the informal economy trade, Cape Town, South Africa. Ecol Soc. 2012;17(2):26. http://dx.doi.org/10.5751/ES-04537-170226.
8. Holmes PM, Rebelo AG, Dorse C, Wood J. Can Cape Town's unique biodiversity be saved? Balancing conservation imperatives and development needs. Ecol Soc. 2012;17(2):28. http://dx.doi.org/10.5751/ES-04552-170228
9. South African National Biodiversity Institute (SANBI). Biodiversity Geographic Information System. Pretoria: SANBI; 2010.
10. South African Department of Environmental Affairs and Tourism (DEAT). Convention on Biological Diversity thematic report on protected areas or areas where special measures need to be taken to conserve biological diversity. Pretoria: DEAT; 2003.
11. Carruthers J. Wilding the farm or farming the wild? The evolution of scientific game ranching in South Africa from the 1960s to the present. Trans R Soc South Afr. 2008;63(2):160–181.
12. Adams WM, Hutton J. People, parks and poverty: Political ecology and biodiversity conservation. Conserv Soc. 2007;5(2):147–183.
13. Sunderland T, Ndoye O, editors. Forest products, livelihoods and conservation. Bogor-Barat: Centre for International Forest Management; 2004.
14. Madubansi M, Shackleton C. Changing energy profiles and consumption patterns following electrification in five rural villages, South Africa. Energy Pol. 2006;34(18):4081–4092. https://doi.org/10.1016/j.enpol.2005.10.011
15. Shackleton CC, Gambiza J, Jones R. Household fuelwood use in small electrified towns of the Makana District, Eastern Cape, South Africa. J Energy South Afr. 2007;18:3.
16. Shackleton S. The informal marula beer traders of Bushbuckridge, Limpopo Province, South Africa. DFID/FRP Winners and Losers in Forest Product Commercialisation, Project No. ZF0140/R7795. Grahamstown: Rhodes University; 2002.
17. Cook B. A bloody trade. London: World Wide Fund for Nature; 2004.
18. Shackleton C, Shackleton S. The *Pterocarpus angolensis* DC. based woodcraft industry in the Bushbuckridge district, South Africa. In: Sunderland T, Ndoye O, editors. Forest products, livelihoods and conservation. Bogor-Barat: Centre for International Forest Management; 2004.
19. Dold AP, Cocks ML. The trade in medicinal plants in the Eastern Cape Province, South Africa. S Afr J Sci. 2002;98(11–12):589–597.
20. Aston Philander L. An ethnobotany of Western Cape Rastafarian bush medicine. J Ethnopharmacol. 2011;138(2):578–594.
21. Cocks ML, Dold AP. The role of 'African Chemists' in the health care system of the Eastern Cape province of South Africa. Soc Sci Med. 2000;51:1505–1515. https://doi.org/10.1016/S0277-9536(00)00050-2
22. Ashforth A. Reflections on spiritual insecurity in a modern African city (Soweto). Afr Stud Rev. 1998;41(3):39–67. https://doi.org/10.2307/525353
23. Natrass N. Who consults sangomas in Khayelitsha? An exploratory quantitative analysis. Working paper. Cape Town: Centre for Social Science Research, University of Cape Town; 1998.
24. Mander M. Marketing of indigenous medicinal plants in South Africa. Rome: Food and Agriculture Organization of the United Nations; 1998.
25. Williams V, Witkowski E, Balkwill K. Volume and financial value of species traded in the medicinal plant markets of Gauteng, South Africa. Int J Sustain Dev World Ecol. 2007;14:584–603. https://doi.org/10.1080/13504500709469757
26. National Environmental Management Act, No 107 of 1998, Government of the Republic of South Africa.
27. Maluleke ML. The Makuleke story [document on the Internet]. No date [cited 2016 Sep 15]. Available from: http://www.earthlore.ca/clients/WPC/English/grfx/sessions/PDFs/session_1/Maluleke.pdf
28. Cousins B. Legislating negotiability: Tenure reform in post-apartheid South Africa. In Juul K, Lund C, editors. Negotiating property in Africa. Portsmouth, NH: Heineman; 2002. p. 67–106.
29. Neimark B. Subverting regulatory protection of 'natural commodities': The *Prunus africana* in Madagascar. Dev Change. 2010;41(5):929–954. https://doi.org/10.1111/j.1467-7660.2010.01666.x
30. General statistical data for City of Cape Town. Cape Town: City of Cape Town Strategic Information and GIS Department; 2007. Unpublished internal report.
31. Statistics South Africa. National census 2011 [document on the Internet]. c2012 [cited 2017 Jan 12]. Available from: https://www.statssa.gov.za/publications/P03014/P030142011.pdf
32. Poswa N, Levy R. Migration study in Monwabisi Park, Khayelitsha – City of Cape Town. Cape Town: Strategic Development Information and GIS Department, Strategic Information Branch; 2006.
33. South African Department of Social Development (DSD). Khayelitsha – Livelihood profile of Khayelitsha and situational analysis of DSD services in the node. Pretoria: DSD; 2007.
34. Rebelo AG. The utilisation of proteas. An unsolicited report for SAPPEX in recognition of its contribution to the Protea Atlas Project. 1996.
35. Cowling RM, Richardson DM. Fynbos: South Africa's unique floral kingdom. Cape Town: Fernwood Press; 1995.
36. Turpie JK, Heydenrych BJ, Lamberth SJ. Economic value of terrestrial and marine biodiversity in the Cape Floristic Region: Implications for defining effective and socially optimal conservation strategies. Biol Conserv. 2003;112:233–251. https://doi.org/10.1016/S0006-3207(02)00398-1
37. Clark BM, Hauck M, Harris JM, Salo K, Russell E. Identification of subsistence fishers, fishing areas, resource use and activities along the South African coast. Afr J Mar Sci. 2002;24:425–437. https://doi.org/10.2989/025776102784528574
38. Steinberg J. The illicit abalone trade in South Africa. Pretoria: Institute for Security Studies; 2005.
39. Loundou P. Medicinal plant trade and opportunities for sustainable management in South Africa [master's thesis]. Stellenbosch: Stellenbosch University; 2008.
40. Nzue APM. Use and conservation status of medicinal plants in the Cape Peninsula, Western Cape Province of South Africa [master's thesis]. Stellenbosch: Stellenbosch University; 2009.
41. Statistics South Africa. Population census South Africa: 2001 [document on the Internet]. c2001 [cited 2017 Jan 12]. Available from: http://www.statssa.gov.za/?page_id=3892

42. Census Plus. Enhanced datasets based on Statistics South Africa Census 2001 [database on the Internet]. c2007 [cited 2017 Jan 12].

43. Finn A, Leibbrandt M, Levinsohn J. Income mobility in South Africa: Evidence from the First Two Waves of the National Income Dynamics Study. SALDRU Working Paper Number 82 / NIDS Discussion Paper 2012/5. Cape Town: SALDRU, University of Cape Town; 2012.

44. Shackleton C. Assessment of the livelihoods importance of forestry, forests and forest products in South Africa. Grahamstown: Rhodes University; 2004. Available from: http://www2.dwaf.gov.za/dwaf/cmsdocs/Elsa/Docs/Forests/Assessment%20of%20the%20Livelihoods%20Importance%20of%20Forestry,%20Forests%20and%20Forest%20Products%20in%20SA,%202004.pdf

45. Shackleton C, Shackleton S. The importance of non-timber forest products in rural livelihood security and as safety nets: A review of evidence from South Africa. S Afr J Sci. 2004;100(11–12):658–664.

46. Hutchings A, Scott A, Lewis G, Cunningham AB. Zulu medicinal plants: An inventory. Pietermaritzburg: University of Natal Press; 1996.

47. Van Wyk BE, Gericke N. People's plants: A guide to useful plants of southern Africa.. Pretoria: Briza; 2000.

48. Cocks ML, Møller V. Use of indigenous and indigenised medicines to enhance personal well-being: A South African case study. Soc Sci Med. 2002;54:387–397. https://doi.org/10.1016/S0277-9536(01)00037-5

49. Ngubane H. Body and mind in Zulu medicine: An ethnography of health and disease in Nyuswa-Zulu thought and practice. London: Academic Press; 1977.

50. Chan K. Chinese medicinal materials and their interface with Western medical concepts. J Ethnopharmacol. 2005;96(1):1–18.

51. Hunt LM, Arar NH, Lakana LL. Herbs, prayer, and insulin. Use of medical and alternative treatments by a group of Mexican American diabetes patients. J Fam Pract. 2000;49(3):216–223.

52. Shahid S, Bleam R, Bessarab D, Thompson SC. 'If you don't believe it, it won't help you': Use of bush medicine in treating cancer among Aboriginal people in Western Australia. J Ethnobiol Ethnomed. 2010;6:18. https://doi.org/10.1186/1746-4269-6-18

53. Weideman M. Tenure reform: The former homelands [document on the Internet]. c2004 [cited 2016 Jul 12]. Available from: http://wiredspace.wits.ac.za/bitstream/handle/10539/275/20_chapter8.pdf?sequence=20

54. United Nations. Workshop on data collection and disaggregation for indigenous peoples; 2004 January 19–21; New York, USA. Available from: http://www.google.co.za/url?sa=t&rct=j&q=jose%20r.%20martinez%20cobo&source=web&cd=1&ved=0CFQQFjAA&url=http%3A%2F%2Fwww.un.org%2Fesa%2Fsocdev%2Funpfii%2Fdocuments%2Fworkshop_data_background.doc&ei=Fcn-T7TiKoyGhQf-6bngDQ&usg=AFQjCNFMb2Ln4AqUEFKk_8ozGuEM8LvcYw&cad=rja

55. Adhikari M. A total extinction confidently hoped for: The destruction of Cape San society under Dutch colonial rule, 1700-1795. J Genocide Res. 2010;12:19–44. https://doi.org/10.1080/14623528.2010.508274

56. Saugestad S. The inconvenient indigenous. Remote area development in Botswana, donor assistance and the first people of the Kalahari [PhD thesis]. Tromsø: University of Tromsø; 1999.

57. Rebelo AG, Holmes PM, Dorse C, Wood J. Impacts of urbanisation in a biodiversity hotspot: Conservation challenges in Metropolitan Cape Town. S Afr J Bot. 2011;77:20–35. https://doi.org/10.1016/j.sajb.2010.04.006

58. Berkes F, Folke C, Gadgil M. Traditional ecological knowledge, biodiversity, resilience and sustainability. Biodivers Conserv. 1995:281–299. https://doi.org/10.1007/978-94-011-0277-3_15

59. Secretariat of the Convention on Biological Biodiversity. Programme of work on protected areas (CBD programmes of work). Montreal: Secretariat of the Convention on Biological Diversity; 2004.

60. Wilshusen P, Brechin S, Fortwrangler C, West P. Reinventing a square wheel: Critique of a resurgent "protection paradigm" in international biodiversity conservation. Soc Nat Resour. 2002;15:17–40. https://doi.org/10.1080/089419202317174002

61. Cocks ML. Biocultural diversity: Moving beyond the realm of 'indigenous' and 'local' people. Hum Ecol. 2006;34:185–200. https://doi.org/10.1007/s10745-006-9013-5

62. Brockington D, Igoe J, Schmidt-Soltau K. Conservation, human rights and poverty reduction. Conserv Biol. 2006;20(1):250–252. https://doi.org/10.1111/j.1523-1739.2006.00335.x

63. Dold AP, Cocks ML. The medicinal use of some weeds, problem and alien plants in the Grahamstown and Peddie Districts of the Eastern Cape, South Africa. S Afr J Sci. 2000;96:467–475.

Externality costs of the coal-fuel cycle: The case of Kusile Power Station

AUTHORS:
Nonophile P. Nkambule[1]
James N. Blignaut[1,2]

AFFILIATIONS:
[1]Department of Economics, University of Pretoria, Pretoria, South Africa
[2]South African Environmental Observation Network, Phalaborwa, South Africa

CORRESPONDENCE TO:
Nonophile Nkambule

EMAIL:
nonoprome@gmail.com

DATES:
Received: 17 June 2016
Revised: 16 Mar. 2017
Accepted: 05 June 2017

KEYWORDS:
power plant; coal mine; plant construction; flue gas desulfurisation

HOW TO CITE:
Nkambule NP, Blignaut JN. Externality costs of the coal-fuel cycle: The case of Kusile Power Station. S Afr J Sci. 2017;113(9/10), Art. #2016-0314, 9 pages. http://dx.doi.org/10.17159/sajs.2017/20160314

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa)

Coal-based electricity is an integral part of daily life in South Africa and globally. However, the use of coal for electricity generation carries a heavy cost for social and ecological systems that goes far beyond the price we pay for electricity. We developed a model based on a system dynamics approach for understanding the measurable and quantifiable coal-fuel cycle burdens and externality costs, over the lifespan of a supercritical coal-fired power station that is fitted with a flue-gas desulfurisation device (i.e. Kusile Power Station). The total coal-fuel cycle externality cost on both the environment and humans over Kusile's lifespan was estimated at ZAR1 449.9 billion to ZAR3 279 billion or 91c/kWh to 205c/kWh sent out (baseline: ZAR2 172.7 billion or 136c/kWh). Accounting for the life-cycle burdens and damages of coal-derived electricity conservatively, doubles to quadruples the price of electricity, making renewable energy sources such as wind and solar attractive alternatives.

Significance:

- The use of coal for electricity generation carries a heavy cost for social and ecological systems that goes far beyond the price we pay for electricity.
- The estimation of social costs is particularly important to the electric sector because of non-differentiation of electricity prices produced from a variety of sources with potentially very dissimilar environmental and human health costs.
- Because all electricity generation technologies are associated with undesirable side effects in their fuel-cycle and lifespan, comprehensive comparative analyses of life-cycle costs of all power generation technologies is indispensable to guide the development of future energy policies in South Africa.

Introduction

Coal is the world's largest single source of energy for electricity generation, fuelling over 40% of global electricity production.[1] In South Africa, about 77% of electricity is derived from this fuel source[2] and recent projections indicate that coal will continue to be the country's primary source of electricity into the distant future. However, the use of coal for energy raises several serious environmental concerns. These concerns include the disruption of large land surface areas and pollution of rivers as a result of coal mining[3]; air pollution, accidents and damage to roadways caused by material inputs transportation[4]; and the warming of the earth as a consequence of greenhouse gas (GHG) emissions from flue stacks[5] (Table 1). The entire chain of coal-based electricity generation is thus associated with dire impacts.[6] However, research has mainly focused on the externalities of the combustion process both internationally[7,8] and locally[9-13]. In addition, the emphasis, even within this phase, is on human health and climate change impacts.[11,13]

While the importance of the externalities associated with the combustion process cannot be understated, additional externalities are associated with upstream and downstream processes.[14] Previous research has therefore led to calls for the consideration of all stages in the coal-fuel chain in order to better inform public policy and private investment.[6] Widening the breadth and width of the studied externalities, as well as embracing the long-term repercussions of energy technologies on environmental and social systems is indispensable in making informed choices on technology selection. This paper aims to advance the understanding of the measurable and quantifiable coal-fuel cycle burdens and costs by assessing these for the Kusile coal-fired power station over its lifespan. A coal-based power and social cost assessment model was developed following a system dynamics approach and was used, among other aims, to address the abovementioned goal.

Externality costs of the coal-fuel chain

Environmental and health impacts in the life cycle of coal have been assessed since 1982 using a range of methods.[16] Two broad categories of methods that have been used by researchers to estimate the externality costs can be identified in the literature – namely, abatement cost methods and damage cost methods. The abatement cost methods use the costs of controlling or mitigating damage as a proxy for the damage caused by an externality. On the other hand, the damage cost methods estimate the actual externality burdens and assign a monetary cost to them using valuation techniques. The damage cost methods can be executed in either a top-down or a bottom-up manner. The top-down approach estimates externality costs of pollutants based on national or regional damages, while the bottom-up approach traces pollutants and other burdens from their initial source, quantifies impacts and monetises impacts using valuation techniques, such as the contingent valuation method (e.g. through directly eliciting willingness-to-pay or willingness-to-accept) or indirect valuation methods (e.g. replacement cost technique or hedonic pricing method). The bottom-up approach is the most preferred approach, but it is data intensive.[17] In most developing countries, primary valuation studies linked to the environmental impacts of energy are lacking. For this reason researchers adjust monetary estimates of externalities from previous studies and transfer them to new contexts.[10] The benefit transfer technique is therefore another method that has been used by researchers.

Table 1: Summary of the coal-fuel cycle environmental and societal impacts

Activity	Biodiversity	Air pollution	GHG emissions	Damage to roads	Accidents	Noise	Water quality
Coal mining impacts							
Coal mining	✓	✓	✓		✓		✓
Beneficiation	✓						✓
Coal transportation	✓	✓	✓	✓	✓	✓	
Plant construction impacts							
Site preparation	✓	✓	✓	✓	✓	✓	✓
Materials production	✓	✓	✓	✓	✓	✓	✓
Materials transportation		✓	✓		✓	✓	
Construction					✓	✓	
Plant operation impacts							
Material inputs production	✓	✓	✓	✓	✓	✓	✓
Material inputs transportation		✓	✓	✓	✓	✓	
Raw material storage: coal, fuels, etc.	✓						✓
Coal combustion	✓	✓	✓		✓	✓	✓
Flue gas clean-up: FGD	✓	✓	✓	✓		✓	✓
Ash and FGD waste disposal	✓	✓	✓	✓			✓

Nkambule[15]

GHG, greenhouse gas; FGD, flue-gas desulfurisation

There are a number of international studies in which attempts have been made to quantify the externality costs of coal-based power generation using various valuation methods. The inflation adjusted externality costs of the reviewed studies are given in Table 2 (2010 values). Table 2 shows that several of these studies were conducted in Europe and the USA, with estimates varying with the country in question, fuel cycle stages studied and the range of impacts investigated. The abatement cost method was the earliest approach that was used by researchers. Most of these early works were focused on air emissions from fuel combustion. The researchers using the top-down damage cost approach, similar to the earlier abatement cost methods, focused on air pollution related impacts from the fuel generation stage. However, the estimates they produced varied with the country in question and the range of impacts studied, with higher estimates for studies in which GHGs were considered in addition to classic air pollutants. The bottom-up approach does not, however, allow for site-specific impacts. With the development of the bottom-up approach, new studies considered site specificity and a few made attempts to consider the entire coal cycle. Yet, the focus was still on GHGs and classic air pollutants. Lastly, the benefit transfer technique has also been used by a number of researchers through transferring and adjusting bottom-up damage cost estimates, for example by Epstein et al.[6] and the International Panel on Climate Change (IPCC)[18]. As expected, they report damage cost estimates that are within the range of estimates reported by studies in which the bottom-up approach was employed.

In South Africa, because of the importance of coal, there are a number of studies that estimated the externality costs linked with coal-based power generation (Table 3). Pretorius[29] and Van Zyl et al.[30] strictly focused on coal mining. Pretorius[29] estimated the water pollution externality cost for Eskom's coal requirements at ZAR0.38/kWh while Van Zyl et al.[30] estimated the impact of coal mining on the quality of water and on climate change (methane) to range from ZAR0.12/t to ZAR0.23/t and from ZAR0.98/t to ZAR6.83/t, respectively. Most of the studies shown in Table 3 focused mainly on the operation phase. Dutkiewicz and De Villiers[31] used the top-down approach to value externalities while the other local studies used the bottom-up approach or benefit transfer technique. Shown in Table 3 are inflation-adjusted costs to 2010 US dollars. As with the international studies, Table 3 shows that the estimates vary with the range of impacts investigated and the fuel cycle stages studied. The estimates produced by Dutkiewicz and De Villiers[31] fall in the lower range of the estimates produced by international studies using a similar approach while those produced by Van Horen[9] are higher than those of Spalding-Fecher and Matibe[10] as a result of a broader range of impacts under consideration. Nonetheless, both these estimates are lower than the damage cost estimates from similar studies conducted abroad, partly because they focus on a subset of the fuel cycle stages.

The rest of the studies in Table 3 are independent studies that were executed in a single project for a specific plant (Kusile), so their externality costs were summed. Nkambule and Blignaut[32] focused on the externalities of mining coal and transporting it to Kusile. They focused on climate change effects, air pollution-related health effects, mortality, morbidity, water pollution, water use externality and the loss of ecosystem services. Riekert and Koch[13], Inglesi-Lotz and Blignaut[12], and Blignaut[11] focused on the coal combustion phase in Kusile, and studied air pollution-related health effects, water consumption externality and climate change effects, respectively. The externality costs of Kusile were approximated to range between 4c/kWh and 26c/kWh – values that are comparable to those produced by similar studies conducted abroad. The outcomes of these four studies are an improvement over the earlier black-box national level studies as they focus on a specific plant and somehow disclose the links between plant performance and environmental or societal burdens. Nonetheless, the studies can also be improved upon by making the cause-effect relationships explicit (through a system dynamics model); by widening the breadth and width of the measurable externality costs within the combustion phase; through assessing indirect burdens, construction phase burdens; and flue-gas desulfurisation (FGD) system burdens; and by embracing the long-term repercussions of the coal-fuel chain on the environmental and social systems.

Research method: A system dynamics approach

The coal-based power and social cost assessment (COALPSCA) model was developed following a system dynamics approach. The model is, in essence, designed to explain the design and performance of a coal-based power plant and its interactions with resource inputs, private costs, externalities, externality costs.

The Vensim software was used to conceptualise, construct, simulate and analyse the model. The model was designed to run for a period of 50 years, in line with the lifespan of Kusile. The model consists of nine sub-models, namely power generation, generation cost, water consumption, water pollution, morbidity and fatalities, ecosystem services loss, air pollution, global pollutants and social cost.

The power generation sub-model models the production of electricity in the Kusile Power Station over its lifespan whereas the generation cost sub-model focuses on the private costs of electricity generation. The rest of the sub-models (with the exception of the social cost sub-model) focus on quantifying and monetising externalities in the coal-fuel chain, so they can be termed the 'externalities sub-models'. The social cost sub-model, on the other hand, integrates all nine sub-models through computing a number of economic and environmental indicators to evaluate coal-based power generation. As the focus of this paper is on

Table 2: International studies on coal-fuel cycle externality costs (2010 values)

Study	Country	Method	Externality cost† (US cents/kWh)	Phases and impacts considered
Schuman and Cavanagh[19]	USA	Abatement	0.14–99.67	Combustion phase (only CO_2 effects)
Chernick and Caverhill[20]	USA	Abatement	7.69–13.62	Combustion phase (air pollution effects, plus GHGs)
Bernow et al.[21]	USA	Abatement	6.61–14.78	Combustion phase (air pollution effects, plus GHGs)
Hohmeyer[22]	Germany	Top-down	0.15–7.82	Combustion phase (air pollution effects, not GHGs)
Ottinger et al.[23]	USA	Top-down	5.80–14.19	Combustion phase (air pollution effects, plus GHGs)
Pearce et al.[24]	UK	Top-down	4.15–22.44	Combustion phase (air pollution effects, plus GHGs)
ORNL and RfF[25]	USA	Bottom-up	0.16–0.71	Mining, transport and combustion phases (air pollution effects, not CO_2)
European Commission[26]	UK	Bottom-up	1.40	Entire fuel chain – including decommissioning (air pollution effects, not CO_2)
	Germany	Bottom-up	3.42	
European Commission[27]	Finland	Bottom-up	0.60–20.59	Entire fuel chain – including decommissioning (air pollution effects, plus GHGs)
	Germany	Bottom-up	2.55–25.53	
	Netherlands	Bottom-up	1.81–26.40	
Epstein et al.[6]	USA	Benefit transfer	9.48 (low)	Mining, transport and combustion phases (air pollution effects, plus GHGs, coal transportation accidents)
			18.07 (best)	
			27.24 (high)	
IPCC[18]	USA	Benefit transfer	7.71	Mining and combustion phases (air pollution effects, plus GHGs)

†Own calculations based on values reported in Sundqvist[17,28]. Values were inflation adjusted (to 2010 values in US cents).

GHGs, greenhouse gases

Table 3: Local studies on external cost of coal-based electricity generation (2010 values)

Study	Method	Externality cost‡ (US cents/kWh)	Phases and impacts considered
Dutkiewicz and de Villiers[31]	Top-down	0.51	
Van Horen[9]	Benefit transfer	0.76–4.27	Mainly combustion phase (air pollution effects, GHGs, water consumption and mining accidents)
Spalding-Fecher and Matibe[10]	Benefit transfer	0.34–2.24	Combustion phase (air pollution effects, GHGs)
Nkambule and Blignaut[32]	Benefit transfer	4.23–25.66	Coal mining and transportation (air pollution effects, GHGs, mortality, morbidity, water use and pollution, etc.)
Inglesi-Lotz and Blignaut[12]	Statistical		Combustion phase (water use externality)
Riekert and Koch[13]	Benefit transfer		Combustion phase (air pollution effects)
Blignaut et al.[33]	Benefit transfer		Combustion phase (CO_2)

‡Own calculations based on values reported in the studies. Values were inflation adjusted (to 2010 values in US cents).

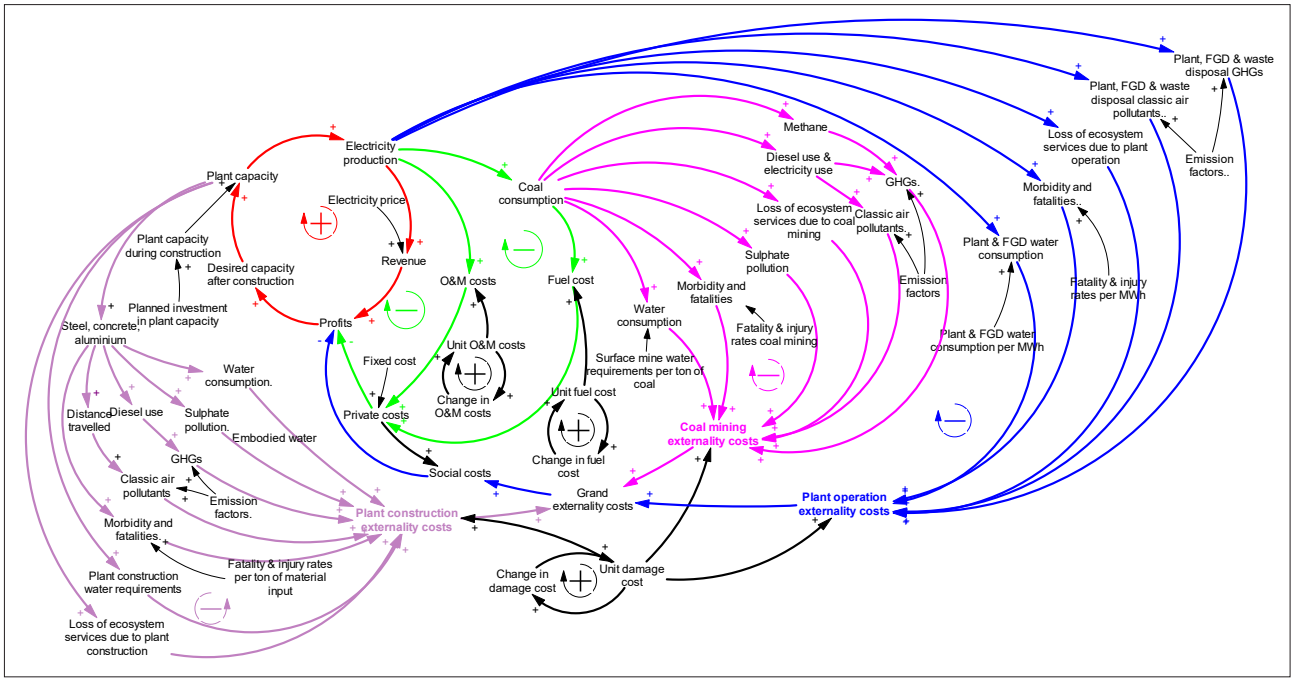
GHGs, greenhouse gases

externalities, only the externalities sub-models and the social cost sub-model are presented and discussed. More discussion on the COALPSCA model is provided in Nkambule[15].

The modelling steps followed in developing the model were: problem formulation, dynamic-hypothesis formulation, model formulation, model validation, and policy design and evaluation. The dynamic hypothesis formulation step involves constructing a working theory that explains the problem. This theory explains and describes the dynamic behaviour of the system premised on the feedbacks and causal structure of the system. The causal loop diagram is a diagram that illustrates in a qualitative manner the linkages and feedback loops of the system and serves as a quick tool for capturing the hypothesis relating to the basis of dynamics. The causal loop diagram displaying the interactions between the key elements and the feedback loops of the modelled system is given in Figure 1. The interactions associated with coal-based power generation, generation cost and externality costs are qualitatively expressed in the causal loop diagram.

System dynamics focuses on understanding the structure of the system so as to provide insight into the behaviour of the system. Accordingly, system dynamics models should include all the important variables that influence a system's behaviour. Table 4 summarises some of the main endogenous, exogenous and excluded variables. The table indicates that many of the key variables were endogenously generated while some exogenous variables also drove the model.

The stock and flow diagrams of the modelled system were constructed and they provide the quantitative relationships between the variables of the system. The stocks or levels are denoted by rectangles and they show accumulations in the system while the flow variables (i.e. inflow and outflow rates) are denoted by valves and they regulate changes in stocks. The stock and flow diagrams of the sub-models are shown in Supplementary figures 1–7 and are explained below.



O&M, operations and maintenance; GHG, greenhouse gas; FGD, flue gas desulfurisation

Figure 1: Causal-loop diagram of the coal-based power and social cost assessment model.

Table 4: Endogenous, exogenous and excluded variables in the system dynamics model

Endogenous variables	Exogenous variables
Gross electricity production	Unit water cost
Net electricity production	Unit coal cost
Operational plant capacity	Unit limestone cost
Coal consumption	Other variable O&M costs
Material inputs inventory (coal, steel, water, diesel, etc.)	Other FGD O&M costs
Pollutant loads (CO_2, SO_2, CH_4, N_2O, etc.)	Growth rate of the various private costs
Dry waste	Escalation of damage costs
Levelised cost of energy	Planned plant capacity
Levelised externality cost	**Excluded variables**
Levelised social cost	
Levelised capital cost	Ecosystem services loss upstream of coal mine
NPV before tax and after tax	Plant construction fatalities and injuries
Social NPV before tax and after tax	Plant construction water pollution
Coal-fuel cycle externality cost of water use	Plant operation water pollution
Coal-fuel cycle fatalities and morbidity costs	Electricity demand

O&M, operations and maintenance; FGD, flue gas desulfurisation; NPV, net present value

Morbidity and fatalities sub-model

The morbidity and fatalities sub-model focuses on injuries and deaths that arise in the coal-fuel cycle. Because accidents are a complicated topic in externality analysis, care needs to be taken to ensure that what is measured are the externality costs. Workers are fully compensated for the risk of accidents to which they are exposed if such cost is fully internalised through the wage rate. However, the high frequency of wage-related strikes in the mining and energy sector in South Africa indicates that workers are not happy with their wages and that their wage rate barely covers an occupational risk premium. In addition, the wage-related strikes coupled with the high level of unemployment in South Africa signify that it is highly unlikely that workers voluntarily bear the occupational risk but rather that they are forced to as they need to provide for themselves and their families.

In addition, a number of serious concerns have been raised with regard to the legislation that governs mining health compensation (i.e. the *Compensation for Occupational Injuries and Diseases Act* and the *Occupational Diseases in Mines and Works Act*), as well as poor service delivery (an insignificant proportion of certified disabled miners receive successful compensation), delays in compensation payment, and virtually no revisions of compensation figures (not even inflationary alterations).[34] It is evident, therefore, that some degree of internalisation is to be expected but the absence of hard data in South Africa with which to approximate and validate the percentage of internalisation meant that we had to base the internalisation risk on the study by the European Commission.[35] The initial unit morbidity and mortality values used in this study (that is, before internalisation) were based on the studies by Van Horen[9], NEEDS[36] and NewExt[37]. Morbidity and mortality values were adjusted with an average of 0% (low), 35% (central) and 50% (high) ranges of internalisation in line with the average assumed internalisation of occupational and non-occupational accidents for non-OECD countries reported in the European Commission[35] study. The internalisation estimates used in the current study, therefore, imply that 50% (low), 65% (central) and 100% (high) estimates for morbidity and mortality are assumed to be externalised.

Supplementary figure 1 represents the structure of the morbidity and fatalities sub-model which consists of two stock variables: unit morbidity value and unit mortality value. Unit morbidity value (UMV, ZAR/person) refers to the value of treating injuries suffered by occupational personnel and the general public. The values for morbidity (low, high and central estimates) were adapted from a study by Van Horen[9] who valued injuries using the cost-of-illness approach in South Africa. The values were adjusted for inflation and some form of internalisation as explained above. The unit value for morbidity (UMV, ZAR/person) is determined by the change in morbidity value (ΔUMV, ZAR/person/year), which is in turn altered by escalation of damage cost (Dmnl/year), which is estimated at the rate of population growth. UMV is mathematically represented by Equation 1:

$$UMV\,(t) = UMV\,(25434) + \int[\Delta UMV]dt, \qquad \text{Equation 1}$$

where UMV (25 434) is the initial value of unit morbidity.

Similarly, the unit mortality value (UMtV, ZAR/person) refers to the economic value for premature mortality. The values for mortality were adapted from the NEEDS[36] and NewExt[37] studies. In transferring estimates from the European Union to the South African context, the context benefit transfer with income adjustment approach was used. Overall, the unit mortality values were adjusted to reflect the disparity of income levels between the European Union and South Africa, and to cater for inflation and some form of internalisation. The unit value for mortality is determined by the change in mortality value (ΔUMtV, ZAR/person/Year), which is in turn altered by escalation of damage cost. The UMtV is represented as:

$$UMtV\,(t) = UMtV\,(245438) + \int[\Delta UMtV]dt. \qquad \text{Equation 2}$$

The unit mortality and morbidity values play a central role in the computation of the coal-fuel cycle fatalities and morbidity costs (CCFMC, ZAR/year). CCFMC is composed of fatalities and morbidity costs streaming from three phases in the coal-fuel cycle – fatalities and morbidity costs from coal mining (FMCM, ZAR/year), from construction (FMC, ZAR/year) and from power generation (FMCPG, ZAR/year) as follows:

$$CCFMC = FMCM + FMC + FMCPG. \qquad \text{Equation 3}$$

The fatality and morbidity costs from all three phases are determined by the deaths and injuries from these phases (which are in essence a function of fatalities and injury rates and the activities occurring in the phases) coupled with the unit mortality or morbidity values, respectively.

Water consumption sub-model

The water consumption sub-model focuses on estimating the coal-fuel cycle externality cost of water use. Estimating the opportunity cost of water use is imperative for a number of reasons. Among these reasons are that water is a scarce resource in South Africa[38]; and that the administered price of water does not reflect the scarcity of water and the price of water seldom reflects the full cost of water delivery[12]. Furthermore, Kusile sits in the Olifants River catchment – a catchment in which water is contested because of the rising water demand from various sectors. The opportunity cost to society of water use when engaging in coal-fired electricity generation was adapted from Inglesi-Lotz and Blignaut[12]. In estimating the opportunity cost, they estimated the shadow price of water when putting water use into coal-fired power generation and into renewable energy technologies. The opportunity cost of water values yielded was, however, adjusted downwards in the current study because the power purchased by the water when put into renewables is in essence not real as these technologies are not yet put into play at such large scales and will not be able to take up the water. The following formula was used to adjust the opportunity cost values:

$$\left[1 - \left(\frac{PS_{SW}}{PS_K}\right)\right] * OC_i,$$

where PS_{SW} is the maximum plant size in MW for solar and wind; PS_K is the maximum plant size in MW of Kusile Power Station and OC_i is the opportunity cost of water with i denoting a low, baseline or high opportunity cost estimate. More details on the adjustment formula can be found in Nkambule[15].

The water consumption sub-model is presented in Supplementary figure 2. The sub-model has one stock variable – the unit opportunity cost of water use (UOCWU, ZAR/m^3) – which plays a pivotal role in the computation of the coal-fuel cycle opportunity cost of water use. The UOCWU is determined by the change in the opportunity cost of water use (ΔOCW, ZAR/m^3/year), which is altered by escalation of damage cost. The UOCWU is given by:

$$UOCWU(t) = UOCWU(1217) + \int[\Delta OCW]dt. \qquad \text{Equation 4}$$

The coal-fuel cycle externality cost of water use (CCExtWU, ZAR/year) is composed of five costs – namely, the opportunity cost of water use in the New Largo colliery during coal mining (OPWCM), construction (OPWC, ZAR/year), power generation (OPWPG), FGD (OPWFGD) and disposal of Kusile's waste (OPWDW) – as follows:

$$CCExtWU = OPWCM + OPWC + OPWPG + OPWFGD + OPWDW. \qquad \text{Equation 5}$$

The opportunity cost of water use during these five processes is in essence functions of the water requirements of the activities occurring in the processes and the unit opportunity cost of water use.

Water pollution sub-model

The water pollution sub-model centres on estimating the coal-fuel cycle water pollution damage cost. Supplementary figure 3 presents the structure of the water pollution sub-model which consists of three stocks: the unit damage cost of sulfate pollution from coal mining, steel production, and aluminium and concrete production. The unit damage costs by these industries represent the damages caused by them on other water users in the eMalahleni catchment, as estimated by Van Zyl et al.[30]

The unit damage cost of sulfate pollution from coal mining (UDSCM, ZAR/ton), steel production (UDSS, ZAR/ton) and aluminium and concrete production (UDSAC, ZAR/ton) is determined by changes in the damage cost of sulfate pollution from coal mining (ΔDSCM, ZAR/ton/year), steel production (ΔDSS, ZAR/ton) and aluminium and concrete production (ΔDSAC, ZAR/ton) which are altered by escalation of damage cost, as follows:

$$UDSCM(t) = UDSCM(0.27) + \int[\Delta DSCM]dt \quad \text{Equation 6}$$

$$UDSS(t) = UDSS(0.79) + \int[\Delta DSS]dt \quad \text{Equation 7}$$

$$UDSAC(t) = UDSAC(0.31) + \int[\Delta DSAC]dt. \quad \text{Equation 8}$$

The coal-fuel cycle water pollution damage cost (CCWPDC, ZAR/year) is composed of two main costs, namely the damage cost of sulfate pollution from coal mining (DCSCM, ZAR/year) and that from Kusile's raw material requirements (DCSMR, ZAR/year). Water pollution damages from the plant operation phase were not considered in the modelling, because Eskom plans to operate the plant under a zero liquid effluent discharge policy. In addition, no major effluents are said to arise from limestone mining and processing[39], so water pollution emanating from such activities was also not quantified. The CCWPDC is represented as:

$$CCWPDC = DCSCM + DCSMR. \quad \text{Equation 9}$$

Ecosystem services loss sub-model

This sub-model is concerned with estimating the coal-fuel cycle cost of lost ecosystem services as a result of siting and operating the power plant and coal mine. These costs are given by the forgone benefits derived from maize farming and ecosystem services generated by grasslands. Supplementary figure 4 presents the structure of this sub-model which consists of two stocks – the unit maize price and unit value of ecosystem services generated by grasslands.

The unit maize price (UMP, ZAR/ton) is an input in the computation of the forgone benefits from maize cultivation. Its initial value was adapted from Blignaut et al.[40] and is determined by the change in maize price (ΔMP, ZAR/ton/year):

$$UMP(t) = UMP(1600) + \int[\Delta MP]dt. \quad \text{Equation 10}$$

The unit value of ecosystem services generated by grasslands (UVEG, ZAR/ha) is an input into the computation of the forgone benefit from ecosystem services generated by grasslands. Its initial value was adapted from Blignaut et al.[40] and is determined by the change in the value of ecosystem goods and services (ΔVEG, ZAR/ha/year) as follows:

$$UVEG(t) = UVEG(510) + \int[\Delta VEG]dt. \quad \text{Equation 11}$$

The coal-fuel cycle cost of lost ecosystem services (CCCLES, ZAR/year) consists of ecosystem services lost as a result of coal mining (ESLCM, ZAR/year) and plant construction and operation (ESLPCO, ZAR/year) and is represented as:

$$CCCLES = ESLCM + ESLPCO. \quad \text{Equation 12}$$

The ecosystem services lost as a result of these two processes are in essence a function of the land areas lost and the unit maize price and unit value of ecosystems generated by grasslands.

Air pollution sub-model

The air pollution sub-model is concerned with estimating the coal-fuel cycle air pollution human health cost. This sub-model structure is presented in Supplementary figure 5 and consists of seven stocks representing the damage cost of the various classic air pollutants studied, namely SO_2, NO_x, particulate matter, nickel, lead, arsenic and chromium.

The coal-fuel cycle air pollution human health cost (CCAPC, ZAR/Year) comprises air pollution health cost from four main processes – coal transportation (CTAC, ZAR/year), plant construction (PCAC, ZAR/year), plant operation (POAC, ZAR/year) and waste disposal (WDAC, ZAR/year) – as follows:

$$CCAPC = CTAC + PCAC + POAC + WDAC. \quad \text{Equation 13}$$

The air pollution health costs from these four processes are in essence functions of transportation distances of coal by road/conveyor, transportation distances of raw material requirements, power production / coal consumption, and electricity use during waste disposal, respectively, coupled with the emission factors of the studied gases and metals and the unit damage cost of these gases and metals (i.e. SO_2, NO_x, particulate matter, arsenic, nickel, lead and chromium).

Global pollutants sub-model

The global pollutants sub-model is concerned with estimating the coal-fuel cycle global warming damage cost. It focuses mainly on three GHGs in the coal-fuel chain, namely CH_4, CO_2 and N_2O. All the studied GHGs and their damages were expressed in their CO_2-equivalence (CO_2e). The structure of this sub-model is presented in Supplementary figure 6 and it contains two stocks, namely the unit damage cost of CO_2 and the unit train emission damage cost. The coal-fuel cycle global warming damage cost (CCGWC, ZAR/year) is composed of global warming damages from four main processes, that is, coal mining and transportation (CMTGWD, ZAR/year), plant construction (PCGWD, ZAR/year), plant operation (POGWD, ZAR/year) and waste disposal (WDGWD, ZAR) as follows:

$$CCGWC = CMTGWD + PCGWD + POGWD + WDGWD. \quad \text{Equation 14}$$

The damages as a result of climate change resulting from the four processes are in essence functions of the various activities occurring in the four processes coupled with emission factors of the studied gases (NO_2, CH_4 and CO_2), global warming potentials of the gases and the unit damage cost of CO_2.

Social cost sub-model

The social cost sub-model is concerned with estimating nine economic indicators, namely levelised externality cost of energy, levelised social cost of energy, cumulative present value revenue, cumulative present value cost, net present value (NPV) before tax, NPV after tax, cumulative present value externality cost, social NPV before tax and social NPV after tax. The structure of the social cost sub-model is presented in Supplementary figure 7 above and is mainly characterised by the indicators. As this paper focuses on externality costs only, the levelised externality cost of energy is the only relevant indicator.

The levelised externality cost of energy (LECOE, ZAR/MWh), is composed of six stocks which reflect the six externalities studied in the coal-fuel cycle, namely cumulative present value (CPV) externality cost of water use, CPV water pollution externality, CPV fatalities and morbidity cost, CPV ecosystem services loss, CPV air pollution cost and CPV global warming damages. All these stocks have more or less similar structures and, to avoid repetition, only the dynamics of the CPV externality cost of water use (CPVExWU, ZAR) is explained. The coal-fuel cycle externality cost of water use together with the present value factor determines the

present value externality cost of water use (PVExWU, ZAR/year), which is an inflow to the CPVExWU, given by:

$$CPVExWU(t) = CPVExWU(0) + \int [PVExWU]dt. \quad \text{Equation 15}$$

The CPVExWU, coupled with cumulative present value net electricity production (PVNEP, MWh), determines the levelised water use externality (LWUEx, ZAR/MWh), as follows:

$$LWUEx = CPVExWU / CPVNEP. \quad \text{Equation 16}$$

The levelised water use externality (LWUEx, ZAR/MWh), together with the levelised values from the remaining stocks, that is, levelised water pollution externality (LWPEx, ZAR/MWh), levelised fatalities and morbidity cost (LFMC, ZAR/MWh), levelised ecosystem services loss (LESSL, ZAR/MWh), levelised air pollution cost (LAPC, ZAR/MWh), and levelised global warming damages (LGWD, ZAR/MWh), are summed to yield the levelised externality cost of energy (LECOE, ZAR/MWh), represented by:

$$LECOE = LWUEx + LWPEx + LFMC + LESSL + LAPC + LFWD \quad \text{Equation 17}$$

Damage cost parameters

The lower, base case and higher range damage cost estimates used in the externalities sub-models are presented in Table 5. The sources of the estimates and how they were adjusted were discussed earlier while describing the various sub-models. More details on the techniques used to derive these estimates are provided in Nkambule[15].

Model outcomes

In this paper, a model was developed for understanding the measurable and quantifiable coal-fuel cycle burdens and externality costs over the lifespan of Kusile Power Station, following a system dynamics approach. A summary of the coal-fuel cycle externalities inventory and selected externality costs outcomes are presented in Tables 6 and 7, respectively. The important findings from this study are summarised below.

The externalities inventory analysis unveiled the plant operation phase as the highest water using phase in the coal-fuel chain (53%), with the FGD system adding about 22% to its baseline water requirements. Water use in the coal mining phase was also found to be high (37%), making the coal-fuel cycle a large yet hidden water user. Another important outcome of the analysis is that the coal mining phase was found to be more prone to injuries than deaths whereas the plant operation phase was found to be more prone to deaths than injuries. Human safety is therefore a serious problem in these two phases. Concerning air pollution loads, CO_2e emissions were estimated at approximately 1583 million tons over the coal-fuel cycle and lifespan of Kusile, with SO_2 emissions at a low 1.7 million tons as a consequence of the installation of the FGD system. Over 85% of the air pollutants emanated from the combustion phase (Table 6).

Table 5: Lower and higher range damage cost estimates versus baseline values

Variable	Units	Lower	Base case	Higher
Unit morbidity value	ZAR/person	9 130	25 434	59 998
Unit mortality value	ZAR/person	69 285	245 438	771 700
Unit opportunity of water use	ZAR/m^3	814	1217	1 619
Unit damage cost of sulfate pollution from coal mining	ZAR/ton	0.19	0.27	0.34
Unit damage cost of sulfate pollution from steel production	ZAR/ton	0.58	0.79	0.99
Unit damage cost of sulfate pollution from aluminium and concrete production	ZAR/ton	0.14	0.31	0.48
Unit damage cost SO_2	ZAR/ton	29 025	51 619	86 778
Unit damage cost NO_x	ZAR/ton	26 735	41 952	64 689
Unit damage cost particulate matter	ZAR/ton	116 739	227 175	402 332
Unit damage cost CO_2	ZAR/ton	104.98	109.89	177.94

Table 6: Coal-fuel cycle externalities inventory

Coal-fuel cycle phase	Water use (Mm^3)	Fatalities (persons)	Morbidity (persons)	Classic air pollutant			CO_2e (Mt)	Land use (ha)
				Mt SO_2	Mt NO_x	Mt PM		
Coal mining/transport	408.2 (36.6%)	49 (9.7%)	716 (77.2%)	(2.5%)	(<1%)	(<1%)	210.5 (13%)	6817
Plant construction	95.5 (8.6%)	<1 (<0.1%)	1 (<0.1%)	(<0.01%)	(<1%)	(<1%)	8.9 (<1%)	1456
Plant operation	342.5 (30.7%)	454 (90.3%)	211 (22.7%)	(97%)	(99%)	(99%)	1348.9 (85%)	
FGD system	248.3 (22.3%)			(<1%)	(<1%)	(<1%)	13.5 (<1%)	
Waste disposal	19.4 (1.7%)			(<1%)	(<1%)	(<1%)	1.2 (<1%)	
Total: Life cycle and lifetime	**1114**	**503**	**928**	**1.7**	**3.9**	**0.4**	**1 582.9**	**8273**

FGD, flue gas desulfurisation; PM, particulate matter

Attaching economic values to the studied externalities yielded a base case total coal-fuel cycle externality cost over Kusile's lifespan of ZAR2172.7 billion or 136c/kWh sent out. The low estimate is ZAR1449.9 billion or 91c/kWh while the high estimate is ZAR3279 billion or close to 205c/kWh sent out (Table 7). The levelised externality costs of energy ranged from ZAR908 to ZAR2052 per MWh (the levelised cost of energy is about ZAR554 per MWh[15]). Accounting conservatively for the life-cycle burdens and damages of coal-derived electricity thus doubles to quadruples the price of electricity. The plant combustion phase with waste disposal housed most of the externality cost (49–107c/kWh), followed by coal mining and transportation (24–61c/kWh) and the FGD system (14–28c/kWh). At the lower end is the construction phase (4–9c/kWh) (Table 7). The combustion phase with its ancillary activities (i.e. FGD system, waste disposal and plant construction) therefore accounts for most of the externality cost (over two-thirds). In addition, the baseline model disclosed that most of the externality cost stems from three types of externalities, namely water use (over 65%), air pollution health cost (over 21%), and climate change effects from GHG emissions (over 10%).

When converted to US cents/kWh the total coal-fuel cycle externality cost ranges between 12c/kWh and 28c/kWh (baseline 19c/kWh), and for the most part falls within the range of the international studies reported in Table 2 and the local studies in Table 3, but is slightly higher because of the inclusion of more externalities and coal-fuel cycle phases. The coal-fuel cycle externality costs estimated in this study are, however, considered lower bound estimates because several externalities were not investigated; those excluded include noise pollution, damages to roads, and some upstream burdens. The true societal and environmental burdens of coal-based power are thus far greater than these numbers suggest.

Way forward for the South African government

The harshest way forward for the South African government in addressing the serious impacts of coal-based electricity, would be to reform the pricing system to properly reflect all the externalities in the price. Reducing the water use externality (over 65%) necessitates policy changes at national and local levels, e.g. requiring power plants to upgrade to dry cooling systems over a reasonable period of time and pricing water well.

Concerning the air pollution related human health effects (over 21%), the government can request retrofits of all existing plants with FGD devices over a reasonable period of time as well as require new plants to be fitted with this device. With regard to climate change effects from GHG emissions (over 10%), the South African government has taken action and intends to internalise the externality cost of carbon emissions on producers of GHGs through a carbon tax of ZAR120/t of CO_2e emissions.[41] The National Treasury[42] has disclosed that introducing the carbon tax will significantly reduce the country's GHGs. In comparison to a business-as-usual scenario, the carbon tax would result in an emissions reduction of 13–14.5% by 2025 and about 26–33% by 2035.

Conclusion

Although coal-based electricity forms an integral part of our day-to-day lives, the use of coal for electricity generation carries a heavy burden for the social and ecological systems that go far beyond the prices we pay for electricity. In this paper, a model was developed based on a system dynamics approach for understanding the measurable and quantifiable coal-fuel cycle burdens and externality costs, over the lifespan of Kusile Power Station. The model showed that accounting for the life-cycle externalities of coal-derived electricity conservatively doubles to quadruples the price of electricity, making renewable energy sources such as wind and solar attractive options. However, because all electricity generation technologies are associated with undesirable side effects, comprehensive comparative analyses of life-cycle costs of all power generation technologies are necessary to guide the development of future energy policies in South Africa.

Acknowledgement

We are grateful to the National Research Foundation (South Africa) for funding.

Authors' contributions

The research was conducted as part of N.P.N.'s PhD, which was supervised by J.N.B.

Table 7: Selected outcomes under low- and high-range damage costs versus baseline

Externality	Units	Lower	Baseline	Higher
Water use	ZAR billion	950.7	1473.5	2142.6
Water pollution		0.2	0.3	0.4
Fatalities and morbidity		0.05	0.2	0.6
Ecosystem loss		6.02	6.1	6.2
Classic air pollutant		268.8	458.2	749.6
Greenhouse gases		224	234.4	379.5
Total		**1449.9**	**2172.7**	**3279.0**
Levelised externality costs of energy	ZAR/MWh	908.0	1370.8	2051.6

Externality cost	Coal mining and transport			Plant operation and waste disposal			FGD system operation			Construction		
	Low	Base	High	Low	Base	High	Low	Base	High	Low	Base	High
c/kWh	24	37	61	49	72	107	14	21	28	4	6	9

FGD, flue gas desulfurisation

References

1. International Energy Agency. Coal [homepage on the Internet]. c2017 [cited 2017 Feb 02]. Available from: http://www.iea.org.za
2. Eskom. Coal power [homepage on the Internet]. c2017 [cited 2017 Jan 28]. Available from: http://www.eskom.co.za
3. Singh G. Mitigating environmental and social impacts of coal mining in India [document on the Internet]. c2008 [cited 2011 Jul 11]. Available from: http://www.ismenvis.nic.in/My_Webs/Digital_Library/GSingh/Mitigating%20Environmental%20and%20Social%20Impacts%20of%20Coal%20Mining%20in%20India.pdf
4. Jorgensen AA. Transport costs and the relevance of externalities [document on the Internet]. c2010 [cited 2011 Jun 25]. Available from: http://www.rra.co.za/?p=16756
5. Odeh NA, Cockeril TT. Life cycle analysis of UK coal fired power plants. Energy Convers Manag. 2008;49:212–220. https://doi.org/10.1016/j.enconman.2007.06.014
6. Epstein PR, Buonocore JJ, Eckerle K, Hendryx M, Stout III BM, Heinberg R, et al. Full cost accounting for the life cycle of coal. Ann N Y Acad Sci. 2011;1219:73–98. https://doi.org/10.1111/j.1749-6632.2010.05890.x
7. Hondo H. Life cycle GHG emission analysis of power generation systems: Japanese case. Energy. 2005;30:2042–2056. https://doi.org/10.1016/j.energy.2004.07.020
8. Zhao Y, Wang S, Duan L, Lei Y, Cao P, Hao J. Primary air pollutant emissions of coal-fired power plants in China: Current status and future prediction. Atmos Environ. 2008;42:8442–8452. https://doi.org/10.1016/j.atmosenv.2008.08.021
9. Van Horen C. Cheap energy – at what cost? Externalities in South Africa's electricity sector. In: Van Horen C, editor. Counting the social costs: electricity and externalities in South Africa. Cape Town: Elan Press and UCT Press; 1997.
10. Spalding-Fecher R, Matibe DK. Electricity and externalities in South Africa. Energy Pol. 2003;31:721–734. https://doi.org/10.1016/S0301-4215(02)00123-4
11. Blignaut JN. Climate change: The opportunity cost of Medupi and Kusile Power Stations. J Energy South Afr. 2012;23(4):67–75.
12. Inglesi-Lotz R, Blignaut JN. Estimating the opportunity cost of water for the Kusile and Medupi coal-fired electricity power plants in South Africa. J Energy South Afr. 2012;23(4):74–84.
13. Riekert J, Koch SF. Projecting the external health costs of a coal-fired power plant: The case of Kusile. J Energy South Afr. 2012;23(4):52–66.
14. Weisser D. A guide to life-cycle greenhouse gas (GHG) emissions from electric supply technologies. Energy. 2007;32:1543–1559. https://doi.org/10.1016/j.energy.2007.01.008
15. Nkambule NP. Measuring the social costs of coal-based electricity generation in South Africa [PhD thesis]. Pretoria: University of Pretoria; 2015. Available from: http://hdl.handle.net/2263/45866
16. Mishra SK. Estimation of externality costs of electricity generation from coal: An OH-MARKAL extension dissertation [PhD thesis]. Columbus, OH: The Ohio State University; 2009. Available from: http://etd.ohiolink.edu/view.cgi/Khadka%20Mishra%20Shruti.pdf?%20osu1259703337
17. Sundqvist T. Electricity externality studies — Do the numbers make sense? [PhD thesis]. Luleå: Luleå University of Technology; 2000. Available from: http://www.diva-portal.org/smash/get/diva2:990535/FULLTEXT01.pdf
18. Parry ML, Canziani OF, Palutikof JP, Van der Linden PJ, Hanson CE, editors. Climate change 2007: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press; 2007.
19. Schuman M, Cavanagh R. A model conservation and electric power plan for the Pacific Northwest. Appendix 2: Environmental Costs. Seattle: NCAC; 1982.
20. Chernick P, Caverhill E. The valuation of externalities from energy production, delivery and use. A report to the Boston Gas Company. Boston, MA: PLC Incorporated; 1989.
21. Bernow S, Biewald B, Marron D. Full-cost dispatch: Incorporating environmental externalities in electric system operation. Electr J. 1991;4(2):20–33. https://doi.org/10.1016/1040-6190(91)90168-S
22. Hohmeyer O. Social costs of energy consumption. Berlin: Springer-Verlag; 1988. https://doi.org/10.1007/978-3-642-83499-8
23. Ottinger RL, Wooley DR, Robinsson NA, Hodas DR, Babb SE. Environmental costs of electricity. New York: Oceana Publications; 1991.
24. Pearce D, Bann C, Georgiou S. The social cost of fuel cycles. Report to the UK Department of Trade and Industry. London: HMSO; 1992.
25. ORNL and RfF. External costs and benefits of fuel cycles. A study for the US Department of Energy and the Commission of the European Communities. Report No. 1-8. Washington: McGraw-Hill/Data Institute; 1994.
26. European Commission. ExternE: Externalities of energy. Vol. 3: Coal and lignite. Luxembourg: Office for Official Publications of the European Communities; 1995.
27. European Commission. ExternE: Externalities of energy. Vol. 7–10. Luxembourg: Office for Official Publications of the European Communities; 1999.
28. Sundqvist T. Regulating externalities in the power sector: Some lessons from previous valuation studies. Miner Energy Raw Mater Rep. 2001;16(1):14–31. https://doi.org/10.1080/14041040119222
29. Pretorius K. Coal mining and combustion: Internalising the cost for a fair climate change debate. Johannesburg: Federation for a Sustainable Environment; 2009.
30. Van Zyl H, Raimondo J, Leiman T. Energy supply sector – coal mining. WWF macroeconomic reforms and sustainable development in South Africa. Johannesburg: Development Bank of South Africa; 2002.
31. Dutkiewicz RK, De Villiers MG. Social cost of electricity production. Report for the National Energy Council. Pretoria: Engineering Research; 1993.
32. Nkambule N, Blignaut JN. The external costs of coal mining: The case of collieries supplying Kusile Power Station. J Energy South Afr. 2012;23(4):85–93.
33. Blignaut J, Koch S, Riekert J, Inglesi-Lotz R, Nkambule N. The external cost of coal-fired power generation: The case of Kusile. Pretoria: Business Enterprises, University of Pretoria; 2011.
34. United States Agency International Development (USAID). Workers' compensation in the Republic of South Africa. c2008 [cited 2014 Aug 13]. Available from: http://ssreform.treasury.gov.za/Publications/Workers'%20Compensation%20in%20the%20Republic%20of%20South%20Africa%20(USAID,%202008).pdf
35. European Commission. ExternE: Externalities of energy, methodology 2005 update [document on the Internet]. c2005 [cited 2011 Mar 17]. Available from: http://ec.europa.eu/research/energy/pdf/kina_en.pdf
36. NEEDS. Final report on the monetary valuation of mortality and morbidity risks from air pollution: Deliverable for WP6 of RS1b of the New Energy Externalities Developments for Sustainability (NEEDS) project [document on the Internet]. c2007 [cited 2011 Mar 12]. Available from: http://www.needs-project.org/RS1b/NEEDS_RS1b_D6.7.pdf
37. NewExt. New Elements for the Assessment of External Costs from Energy Technologies [document on the Internet]. c2004 [cited 2011 Mar 19]. Available from: http://www.ier.uni-stuttgart.de/forschung/projektwebsites/newext/
38. Turton A. Three strategic water quality challenges that decision-makers need to know about and how the CSIR should respond [document on the Internet]. c2008 [cited 2011 Jun 25]. Available from: https://cdn.mg.co.za/uploads/keynoteaddresscsir2008.pdf
39. BCS-Incorporated. Energy and Environmental Profile of the U.S. Mining Industry [document on the Internet]. c2002 [cited 2011 Nov 17]. Available from: http://www1.eere.energy.gov/
40. Blignaut J, Mander M, Schulze R, Horan M, Dickens C, Pringle K, et al. Restoring and managing natural capital towards fostering economic development: Evidence from the Drakensberg, South Africa. Ecol Econ. 2010;69:1313–1323. https://doi.org/10.1016/j.ecolecon.2010.01.007
41. National Treasury. Carbon tax policy paper: Reducing greenhouse gas emissions and facilitating the transition to a green economy [document on the Internet]. c2013 [cited 2013 Dec 10]. Available from: http://www.treasury.gov.za
42. National Treasury. Media Statement: Publication of the carbon tax modelling report [document on the Internet]. c2016 [cited 2017 Jan 27]. Available from: http://www.treasury.gov.za

Exploring the relationship between entry requirements and throughput rates for honours students

AUTHOR:
Mike Murray[1]

AFFILIATION:
[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

CORRESPONDENCE TO:
Mike Murray

EMAIL:
murraym@ukzn.ac.za

DATES:
Received: 15 Sep. 2016
Revised: 18 Jan. 2017
Accepted: 09 June 2017

KEYWORDS:
Heckman model; sample selection bias; regression model

HOW TO CITE:
Murray M. Exploring the relationship between entry requirements and throughput rates for honours students. S Afr J Sci. 2017;113(9/10), Art. #2016-0281, 6 pages. http://dx.doi.org/10.17159/sajs.2017/20160281

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

In order for a student to enrol in an honours programme at the University of KwaZulu-Natal (UKZN), a weighted average mark for their final year of undergraduate study must exceed a particular threshold value. Students are then ranked according to this weighted average mark, with entry into the honours programme offered on a top-down basis, within the constraints of teaching resources and space. A proposal has been made at UKZN to remove existing barriers for entry into an honours programme, i.e. to allow entry to any student who has completed a 3-year undergraduate degree with a major in that discipline. The impact of such a decision was investigated. By lowering the requirement for entry into an honours programme, one is expected to predict how a new cohort of students will perform. Apart from obviously having a lower weighted average mark for their final year of undergraduate study, these new students may also differ in other unobservable ways which need to be accounted for. In a regression modelling context, one is asked to predict outside the range of a collected data set. A Heckman selection model was used to account for a possible self-selection bias that may arise because the subpopulation for which a prediction is required (namely those new students who will now be able to enter an honours programme), may be significantly different from the population of UKZN undergraduate students who are currently permitted entry to an honours programme.

Significance:

- A modelling technique that accounts for a possible sample selection bias was used to determine the impact of lowering the entry requirements into the honours programme at UKZN to allow entry to any student who has completed a 3-year undergraduate degree.

Introduction

In order to prepare students for entry into a job market that is rapidly evolving, 3-year undergraduate degrees that were once common in North America and parts of East Asia and South America have been replaced with 4-year undergraduate degrees. In Australia, the University of Sydney recently announced a radical reorganisation of their course offerings with an emphasis on the restructuring of their 3-year degrees into 4-year degrees. Noting that only a 'relatively small fraction' of students left their university after completing a 3-year degree, by embedding an honours degree within a 4-year undergraduate degree not only would the university be able to better prepare students for postgraduate study but they would also be able to substantially reduce the number of degrees that they have to offer. In South Africa, a similar debate is taking place at the University of KwaZulu-Natal (UKZN). In some disciplines, such as Engineering, a 4-year qualification has for many years been the norm. In the College of Management Sciences, students have the option to complete a 3-year undergraduate degree (which is then followed by an additional year of honours study) or a 4-year degree. When faced with this choice, students in that college seem to prefer enrolling for a 4-year degree.

This paper is concerned primarily with the identification of the impact of the decision to lower the requirement for entry into honours on the throughput rate within that programme. The positive (or negative) consequences of allowing these (possibly weaker) students to enrol for an honours degree need to be weighed against the 'massification' of undergraduate programmes that is taking place, in which an increasing number of undergraduate students are wanting to improve their qualifications by enrolling in an honours programme.

Given that there is a financial cost incurred in training these students, by relaxing the entry requirement one may simply be increasing the number of students who are being set up to fail. A White Paper outlining a National Plan for Higher Education[1] made mention of this problem, highlighting at an undergraduate level the need for a planned expansion of the system to be 'responsibly managed and balanced in terms of the demand for access, the need for redress and diversification, the human resource requirements...and the limits of affordability and sustainability'. The document further states[1]:

> *While it is possible to achieve rapid enrolment growth without extra expenditure, the penalties for doing so are harsh. Experience...shows that expansion without new investment results in overcrowded facilities, low morale of academic staff, poor quality programmes, a fall in research output and quality, and...devalued products of higher education.*

More recently[2] the debate around curriculum change has grown; the Council on Higher Education released a document in 2013 proposing an overhaul of the undergraduate curriculum in South Africa. However, the focus of this paper is on the consequences of streamlining the pipeline for entry into honours.

Undeniably, being able to increase the number and knowledge base of students in our higher education institutions has many benefits. Research into academic performance at an undergraduate level has traditionally identified prior performance as being an important predictor variable. However, more recent studies[3-9] have found that academic success also depends significantly on the perception of a student regarding their learning environment and how

successful they have been in coping with difficult situations they have encountered at university[10-12]. Economic, social and cultural factors also have an important role to play in determining academic performance at an undergraduate level. However, the socio-economic factors that affected one's performance at an undergraduate level (in the above studies[3-12]) may not be as strongly pronounced for an honours degree. Nevertheless, including these factors as covariates in our prediction model will be important for the analysis that follows.

Currently, entry into an honours programme at UKZN is not automatic. A weighted average mark in the final year of undergraduate study must exceed a threshold value of 55% before a student is allowed to enrol for an honours degree. Will the lowering of this entry requirement to a weighted average mark of 50% or more have a serious impact on the throughput rate that will then be recorded by students entering a fourth year (honours) study at UKZN?

International research on this matter is usually restricted to cases in which a group of students for which the relaxation of entry requirements does not apply are compared with another group for which a relaxation of standards has been applied.[13-16] One can then look for a difference in performance between these two cohorts – controlling, where necessary, for background variables that may also impact on their performance in their honours year of study. In this study, however, it was not possible to observe an honours-based performance for students who, under the current criteria, do not qualify for entry into honours. Consequently, the focus in this paper is different: to determine how this new cohort of students would perform if they had been given the opportunity to enrol for an honours degree at UKZN. A Heckman model needs to be used to adjust for a possible self-selection bias that may arise because the outcome of interest (in our case a suitably chosen measure of performance in the honours programme) can be observed for only a subset of students who were previously eligible for entry into honours.

The Heckman model and its use in a sociological and economic setting has been well documented in the literature.[17-21] In an educational setting, however, its application has been restricted mainly to identifying what sort of causal effect a particular level of education has on a given socio-economic response variable (such as earnings). In fact, a detailed review of the literature[22] has found that only 14% of 386 articles discussing the problem of selection bias have done so in an educational context with a Heckman selection model then fitted to the collected data. The application of a Heckman model to our problem – for which 'entry into honours' replaces 'level of education' as the treatment variable and a weighted average mark in honours is used as a response variable – could well be novel, but is most certainly well supported by a number of other applications in the literature.

An analysis based on a weighted average mark in honours

This study focuses on a very specific cohort of students: those who completed an undergraduate degree at UKZN during the period 2011–2013. This cohort comprises 9398 students who obtained a weighted average mark in their final year of study that exceeded a threshold value of 50%. Under the prevailing set of rules allowing for a progression into honours, 3233 students from this sample were allowed to enrol for an honours degree because their weighted average mark exceeded a threshold value of 55%. An additional 2021 students from other universities were also allowed to enrol. Apart from noting very briefly that these other students generally did not perform as well in the honours programme as their counterparts who completed undergraduate studies at UKZN, the focus of this paper will rest entirely on comparing the observed performance (in honours) of the 3233 students who completed undergraduate studies at UKZN and were able to enter honours, with the performance that one would expect from a hypothetical cohort of 9398 UKZN graduates if the other 6165 students who achieved a weighted average mark in their final year of study of 50–55% joined the 3233 students currently in the honours programme.

To measure their performance in honours, each course was assigned an appropriate credit point value; an honours degree is considered to be completed when a certain number of credit points have been acquired. Using these credit point values as appropriate weights, a weighted average mark for all the compulsory courses was used as response variable Y_i. For example, if eight courses must be passed in order to complete an honours programme and each course has been given a credit weighting of 16 points, then

$$\frac{16 * \textit{Mark for course } 1 + \ldots + 16 * \textit{Mark for course } 8}{16 + \ldots + 16}$$

will be the observed weighted average mark in honours for that student. If a course had to be repeated then both marks together with their appropriate weights were included in the response variable calculation for that student. Permission to use this data set was given by the Teaching and Learning Office at UKZN.

The results in Table 1 suggest that students who completed an undergraduate degree at UKZN performed significantly better in the UKZN honours programme than those who completed undergraduate studies at another university (t=12.185). Perhaps a significant number of the students who transferred from another university were weaker students who were refused entry into an honours programme at that university. There may also be other reasons for such an anomaly. For example, the demographic breakdown shown in Table 2 indicates that the students from other universities were older than those who entered from a UKZN undergraduate programme.

Table 1: Weighted average marks (%) achieved by students enrolled in a honours course at the University of KwaZulu-Natal (UKZN) over the period 2011–2013

Undergraduate alma mater	Number of students	Mean	Standard deviation
UKZN	3233	63.97	11.55
Non-UKZN	2021	58.39	18.45

Table 2: Student enrolment figures cross-tabulated by undergraduate background

Covariate	UKZN alumni	Non-UKZN alumni
Received honours funding		
Yes	693	178
No	2540	1843
Sex		
Female	1990	1222
Male	1243	799
Race		
Black	1806	1427
White	347	180
Indian	981	358
Other	99	56
Age (years)		
Under 20	10	2
20–24	2921	800
25–29	241	489
30–34	30	236
35+	31	494

Because the goal was to assess the impact of lowering the entry requirement on the throughput rate in the honours programme, the actual performance of the 3233 UKZN graduates enrolled in honours was compared with the expected performance of all 9398 graduates who would have been allowed to enrol if the entry requirement had been a weighted average mark of 50% or more.

Table 3 shows the results of regressing the weighted average mark in honours of each of the 3233 students against a set of appropriately chosen predictor variables. Ordinary least squares methods were used to produce the parameter estimates that appear in Table 3.

Table 3: Parameter estimates associated with a regression model fitted to their weighted average mark Y_i in honours

Predictor variable	Estimate	Standard deviation	95% Confidence interval
Constant intercept	40.9	2.60	[35.87; 46.09]
Weighted average mark in final year of undergraduate study (%)	0.52	0.03	[0.47; 0.57]
Age (years)	-0.17	0.07	[-0.32; -0.02]
Matric point score	-0.016	0.02	[-0.05; 0.02]
Female (0/1)	1.56	0.38	[0.81; 2.32]
Black (0/1)	-4.73	0.43	[-5.58; -3.88]
Honours funding (0/1)	1.40	0.48	[0.45; 2.36]
College			
Agriculture, Engineering and Science (0/1)	0.65	0.51	[-0.35; 1.65]
Health Sciences (0/1)	-3.14	0.97	[-5.05; -1.23]
Humanities (0/1)	-1.72	0.49	[-2.69; -0.76]

Law and Management is the baseline category for the Colleges.

For the interpretation of Table 3, 'female' refers to a 0/1 indicator variable that is set equal to 1 for a female student. 'Black' refers to a 0/1 indicator variable that is set equal to 1 for a black student. Similarly, 'honours funding' refers to a 0/1 indicator variable that is set equal to 1 for a student who has managed to obtain some form of external funding for honours study. Because UKZN is composed of four colleges (College of Agriculture, Engineering and Science, College of Health Sciences, College of Humanities and College of Law and Management Studies), indicator variables for the first three colleges were included as covariates in Table 3 with an effect for the College of Law and Management Studies forming part of the intercept term in the above model structure. These covariates were chosen based on previous research[23-26] at UKZN that has identified them as being important predictor variables for determining performance at UKZN.

As one would expect, the weighted average mark that a UKZN undergraduate student obtains in their final year of undergraduate study correlates strongly and positively with the weighted average mark that they record for their honours year of study. Older students tend to not perform as well as their younger counterparts and female students perform better than their male counterparts. Having some form of honours funding also has a positive effect on performance in honours. Significant college effects are also present, with students from the College of Health Sciences scoring 3.139% lower than that of a baseline student from the College of Law and Management Studies. Students in the College of Agriculture, Engineering and Science score on average 0.649% more than a baseline student from the College of Law and Management Studies. The most significant effect, however, appears to be associated with race, with black students scoring on average 4.733% lower than their counterparts from other race groups.

In the above regression model setting, a set of parameter estimates $\{\beta_0, \beta_1, \ldots, \beta_p\}$ is identified that optimally links a set of demographic variables $\{X_{1i}, \ldots, X_{pi}\}$ to the weighted average mark Y_i that a student i obtains in their honours year with e_i forming an error term accounting for a lack of fit in the following model structure:

$$Y_i = \beta_0 + X_{1i}\beta_1 + \ldots + X_{pi}\beta_p + e_i. \qquad \text{Equation 1}$$

The results given in Table 3, however, relate to a very specific data set – namely those 3233 students with an undergraduate degree from UKZN who were allowed to enrol for an honours degree under the current set of rules which the university is now wanting to relax. With these results in hand, a student with a demographic profile of, for example:

- Weighted average mark in final year of undergraduate study = 65%;
- Age (years) = 25;
- Matric point score = 43;
- Female = 1;
- Black = 1;
- Honours funding = 0;
- College:
  - Agriculture, Engineering and Science = 1
  - Health Sciences = 0
  - Humanities = 0

could be expected to record (on average) the following weighted mark for their honours degree:

$$E(Y) = 40.9 + 0.52(65) - 0.17(25) - 0.016(43) + 1.56(1) - 4.73(1) + 0.65 = 67.24.$$

Because these results are estimated using a sample of 3233 UKZN graduates who were able to enrol in an honours programme, one cannot simply assume that the same sort of effects would occur if UKZN were to lower their entry requirement to include UKZN graduates with a weighted average value for their final year of undergraduate study of 50–55%. Apart from obtaining a lower weighted average score, these students may also differ in other important respects from the 3233 UKZN graduates who are currently eligible for entry into honours. To correct for a possible (sample selection) bias that may arise if we were to simply assume that the same set of results that appear in Table 3 would apply to this new cohort, a Heckman selection model was fitted to a now extended set of data comprising the 6165 UKZN graduates with a weighted average score for their final year of undergraduate study of 50–55% and the 3233 UKZN graduates currently in the honours programme for which we have a weighted average score for honours.

A Heckman selection model attempts to solve the problem associated with a potential sample selection bias by linking the actual selection process for entry into an honours programme to a set of appropriately chosen predictor variables $\{Z_{1i}, \ldots, Z_{ki}\}$. Let R_i denote a 0/1 indicator variable that models the selection process that is actually taking place. More specifically, assume entry into an honours programme (i.e. setting $R_i = 1$) is determined by the following decision rule:

$$\text{Set } R_i = 1 \text{ if } \theta_0 + \theta_1 Z_{1i} + \ldots + \theta_k Z_{ki} + u_i > 0, \qquad \text{Equation 2}$$
$$\text{else set } R_i = 0,$$

where $u_i \sim N(0,1)$ denotes a random error term used to reflect the fact that once the threshold value for admission into an honours programme has been passed, admission into an honours programme is not necessarily automatic. For example, UKZN applicants are ranked according to the

weighted average mark that they record in their final year of undergraduate study with 'top-slicing' then occurring based on other factors relating to classroom size and staff supervisory constraints. Parameter estimates for Equation 2 are derived from the extended set of data that include the 6165 UKZN graduates with a weighted average score for their final year of undergraduate study of 50–55% and the 3233 UKZN based undergraduate students who were allowed to enrol in honours.

As a second stage in the Heckman model, a regression model is then fitted to the weighted average marks Y_i of those 3233 UKZN undergraduate students who **were able to enrol** for an honours degree at UKZN:

$$Y_i = \beta_0 + X_{1i}\beta_1 + ... + X_{pi}\beta_p + e_i. \qquad \text{Equation 3}$$

To complete the model formulation, u_i and e_i are assumed to have a bivariate normal distribution with

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} | X_i, Z_i \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & p\sigma \\ p\sigma & 1 \end{pmatrix} \right).$$

In the context of this paper, it is important to note that one is only observing an honours performance based outcome for someone who has actually been selected for entry into that programme. For the self-selected subpopulation of random variables Y_i that we are able to observe, Heckman has shown that[5,7,8]:

$$E[Y_i | X_i, Z_i, R_i = 1] = X_i\beta + E[e_i | u_i \geq -Z_i] = X_i\beta + p\sigma\lambda_i$$

and

$$\mathrm{Var}[Y_i | X_i, Z_i, R_i = 1] = \sigma^2(1-p^2\delta_i)\ \delta_i = \lambda_i(\lambda_i + Z_i\delta),$$

where

$$\lambda_i = \frac{\phi(Z_i\delta)}{\Phi(Z_i\delta)}$$

denotes an inverse Mills ratio, ϕ a standard normal density and Φ a standard normal cumulative distribution function.

One could estimate the unknown model parameters in Equations 2 and 3 simultaneously using a maximum likelihood method based on the joint normality assumption of u_i and e_i. As an alternative, however, if λ_i were observable, then ordinary least squares could be used on

$$Y_i = X_i\beta + p\sigma\lambda_i + w_i, \qquad \text{Equation 4}$$

with w_i denoting a zero mean error term with variance $\sigma^2(1-p^2\delta_i)$ that is now distributed independently of $\{X_i, Z_i\}$. As λ_i is unknown, Heckman proposed the following two-step procedure to estimate β in the outcome Equation 3:

- Step 1: Use all 9398 observations on Z_i that are available, namely those associated with students who were eligible for entry into honours (i.e. those with $R_i = 1$ and also those who were refused entry into honours (i.e. those with $R_i = 0$ to generate a maximum likelihood estimator $\hat{\delta}$ for δ based on the selection Equation 2. This estimate can then be used to compute the following inverse Mills ratio term for each observation; viz.
  $$\hat{\lambda}_i = \frac{\phi(Z_i\hat{\delta})}{\Phi(Z_i\hat{\delta})}$$
- Step 2: Now use the self-selected sample only (i.e. all 3233 observations for which we have $R_i = 1$) to run an ordinary least squares regression on the following equation that is implied by the formulation given by Equation 4:
  $$Y_i = X_i\beta + \beta_\lambda\lambda_i + w_i;\ w_i \sim N(0, \sigma_w^2)\ \beta_\lambda = p\sigma$$

Because

$$\mathrm{Var}[w_i | X_i, Z_i, R_i = 1, \hat{\delta}] = \sigma^2\,(1-p^2\delta_i),$$

an appropriate correction needs to be made to the variance covariance matrix that is associated with $(\hat{\beta}, \hat{\beta}_\lambda)$.

Tables 4 and 5 present the parameter estimates that result from fitting a Heckman model to the UKZN undergraduate based data set that is available. Table 6 contains a parameter estimate for p with a likelihood ratio based Wald test for $p=0$ rejected at all of the major levels of significance (chi-square=23.55, $p\approx 0$). This result confirms that the results in Table 3 do in fact need to be corrected for a sample selection bias because unobservable factors making up the error term u_i in the selection equation are correlating positively with the error term e_i that forms part of the response model determining the weighted average mark for honours. One cannot simply extrapolate the results shown in Table 3 to the new cohort of students and assume that these exact same results will be true for this new cohort of students.

Table 4: Parameter estimates associated with fitting the Heckman model given in Equation 3

	Estimate	Standard deviation	95% Confidence interval
Constant	34.46	3.20	[28.19; 40.73]
Weighted average mark in final year of undergraduate study	0.59	0.03	[0.52; 0.67]
Age (years)	-0.28	0.09	[-0.48; -0.09]
Matric point score	-0.03	0.01	[-0.06; 0.00]
Female	1.60	0.39	[0.82; 2.38]
Black	-4.76	0.41	[-5.57; -3.93]
Honours funding	1.45	0.51	[0.43; 2.47]
College			
Agriculture, Engineering and Science	0.62	0.46	[-0.27; 1.52]
Health Sciences	-2.64	0.94	[-4.50; -0.78]
Humanities	-1.07	0.50	[-2.04; -0.086]

Law and Management is the baseline category for the Colleges.

Table 5: Parameter estimates associated with the selection model given in Equation 2

	Estimate	Standard deviation	95% Confidence interval
Constant	-2.23	0.19	[-2.61; -1.84]
Weighted average mark in final year of undergraduate study	0.05	0.002	[0.04; 0.05]
Years registered as an undergraduate student	-0.14	0.01	[-0.17; -0.11]
Age (years)	-0.04	0.01	[-0.05; -0.03]
Black	0.11	0.04	[0.04; 0.18]
Undergraduate funding	0.34	0.03	[0.27; 0.41]

Table 6: Parameter estimate associated with the correlation structure between e_i and u_i

	Estimate	Standard deviation	95% Confidence interval
ρ	0.168	0.034	[0.101; 0.234]

Replacing the '3-year plus honours' degree structure with a single 4-year structure is synonymous with lowering the honours entry requirement for UKZN undergraduates to 50% and excluding all foreign (non-UKZN applicants) who want to do an honours degree at UKZN. For example, for a 25-year-old, black, female UKZN undergraduate (with 43 matric points) who obtained a weighted average mark of 50% for her final year of undergraduate study and wants to enrol for an honours degree in the Faculty of Humanities, for which she did not receive any extra funding, the parameter estimates that appear in Table 4 indicate that she would on average record the following weighted average mark for honours:

$$E(Y) = 34.46 + 0.59(50) - 0.28(25) - 0.03(43) + 1.60(1) - 4.76(1) + 1.45(0) - 1.07(1) = 51.44.$$

If one were to make use of the results that appear in Table 3 which do not correct for a sample selection bias, then a much higher expected mark would erroneously be associated with the performance of this student in honours:

$$E(Y) = 40.98 + 0.52(50) - 0.17(25) - 0.016(43) - 4.73(1) + 1.56(1) - 1.72(1) = 57.15.$$

One cannot therefore use a regression model to naively extrapolate beyond the range of the data and expect to obtain appropriate results when attempting to answer the question posed in this paper. Furthermore, significant college based effects are recorded in Table 4 which suggest that any decision to relax the entry requirement for a specific college should not be applied as a blanket rule for all colleges. For example, if a weighted average mark of 50% is all that is required to complete an honours year of study, then the results in Table 4 suggest that students in the College of Health Sciences, black students, older students and male students do not perform as well as their counterparts in their honours year of study. Having access to some form of funding seems to improve results, but it is important to note that we are dealing with an associative rather than causative effect in this analysis. For example, funding may be associated with better results because higher achievers are more likely to receive funding; thus they perform better because they are higher achievers rather than because of the funding they received.

In conclusion, the purpose of this paper was twofold. Firstly, to contribute to a debate on the restructuring of undergraduate degrees in which there is a danger associated with naively extrapolating a regression model beyond the range of the data. Secondly, to provide a modelling technique that accounts for a possible sample selection bias that may arise because the profile of the proposed students may be very different from the profile of the students in the sample from which these inferences were drawn. To answer more specifically the question posed: lowering the requirement for entry into honours for UKZN graduates will result in this new cohort not performing as well in honours as their counterparts who currently are allowed into honours. However, the results in Table 1 indicate that students from other universities who currently are allowed into honours do not perform as well as their UKZN graduate counterparts. Therefore, replacing some of these students with a new cohort of UKZN graduate applicants may in fact improve throughput rates in some of the Colleges at UKZN.

Possible limitations

Because this study is primarily exploratory in nature, there are several issues or limitations. Because the focus was on UKZN data only, results from this study cannot necessarily be extrapolated to other universities in South Africa. However, showing how a self-selection bias can be corrected for using a Heckman model can easily be extended to data from another university, once the need to correct for a possible self-selection bias has been fully appreciated.

Because UKZN wants to relax the entry requirement for honours study from a weighted average mark of at least 55% to that of 50% for the final year of undergraduate study, this study focused on a very specific cohort of new students, namely those in the 50–55% band who would then be eligible to join students in the 55+% bracket who currently are allowed to enrol in an honours programme. If UKZN were to further relax the requirements, for example, to include students in the 45–55% band, then a very different set of results may arise.

Given that the aim was primarily to relate the weighted average mark recorded in honours to the mark obtained in the third year of undergraduate study, the inclusion of additional factors related to race, gender and matric point score may be viewed as complicating the analysis unnecessarily. The estimates that appear in Tables 3 and 4, however, will be appropriate only if the error term e_i that is referenced in Equations 1 and 3 is no longer correlated with any of the variables $\{X1_i \ldots X_{pi}\}$ that are included as predictors in these respective regression model structures. By including enough of these variables in the model building process, the hope is to reduce the omitted variable bias problem that might arise if these confounders in the analysis are not controlled for.

References

1. South African Department of Education (DoE). National plan for higher education. White Paper 3 section 2.1.1. Pretoria: DoE; 1997.
2. South African Council on Higher Education (CHE). Report of the task team on undergraduate curriculum structure: A proposal for undergraduate curriculum reform in South Africa. Pretoria: CHE; 2013.
3. Astin AW. Student involvement: A developmental theory for higher education. J Coll Student Dev.1999;40:518–529.
4. Burton LJ, Dowling D. In search of the key factors that influence student success at university. Paper presented at: 28th HERDSA Annual Conference: Higher Education in a Changing World (HERDSA 2005); 2005 July 3–6; Sydney, Australia. Available from: https://eprints.usq.edu.au/753/1/Burton_Dowling_HERDSA2005_PV.pdf
5. Entwistle NJ, Wilson JD. Degrees of excellence: The academic achievement game. London: Hazell Watson and Viney Ltd; 1977.
6. Dayioglu M, Serap TA. Gender differences in academic performance in a large public university in Turkey. High Educ. 2007;53:255–277.
7. Bowden J, Marton F. The university of learning: Beyond quality and competence. Abingdon-on-Thames: Routledge Falmer, 1998.
8. Jonassen DH, Grabowski BL. Handbook on individual differences, learning and instruction. Mahwah, NJ: Lawrence Erlbaum Associates;1993.
9. Pustjens H, Van De Gaer E, Van Damme J, Onghena P. Effects of secondary school on academic choice and on success in higher education. Sch Eff Sch Improv. 2004;15:281–311. https://doi.org/10.1080/09243450512331383222
10. Lizzo A, Wilson K, Simons R. University students perceptions of the learning environment and academic outcomes: Implications for theory and practice. Stud High Educ. 2002;27:27–52. https://doi.org/10.1080/03075070120099359
11. Drew PY, Watkins D. Affective variables, learning approaches and academic achievement: A causal modelling investigation with Hong Kong tertiary students. Brit J Psychol. 1998;68:173–188. https://doi.org/10.1111/j.2044-8279.1998.tb01282.x
12. Zajacova A, Lynch SM, Espenshade TJ. Self-efficacy, stress and success in college. Res High Educ. 2005;46(6):677–706. https://doi.org/10.1007/s11162-004-4139-z
13. Sternberg RJ, Bonney C, Gabora L, Merrifield M. WICS: A model for college and university admissions. Educ Psychol. 2012;47(1):30–41. https://doi.org/10.1080/00461520.2011.638882
14. Atkinson R, Geiser S. Reflections on a century of college admissions tests. In: Soares JA, editor. SAT wars: The case for test-optional college admissions. New York: Teachers College Press; 2012. p. 23–49.
15. Soares JA. The future of college admissions: Discussion. Educ Psychol. 2012;47(1):66–70. https://doi.org/10.1080/00461520.2011.638902

16. Niu SX. Minority student academic performance under the uniform admission law: Evidence from the University of Texas at Austin. Educ Eval Pol Anal. 2010;32(1):44–69. https://doi.org/10.3102/0162373709360063

17. Heckman JJ, Lalonde R, Smith J. The economics and econometrics of active labor markets programs. In: Ashenfelter A, Card D, editors. Handbook of labor economics. Vol. 3. New York: Elsevier; 2000.

18. Schwiebert J. Estimation and interpretation of a Heckman selection model with endogenous covariates. Emp Econ. 2015;49:675–703. https://doi.org/10.1007/s00181-014-0881-z

19. Heckman JJ. Dummy endogenous variables in a simultaneous equation system. Econometrica.1978;46:931–959. https://doi.org/10.2307/1909757

20. Heckman JJ. Sample selection bias as a specification error. Econometrica. 1979;47:153–161. https://doi.org/10.2307/1912352

21. Briggs D. Causal inference and the Heckman model. J Educ Behav Stat. 2004;29(4):397–420. https://doi.org/10.3102/10769986029004397

22. Deschacht N, Goeman K. Selection bias in educational issues and the use of Heckman's sample selection model. In: Contemporary Economic Perspectives in Education. Leuven: Leuven University Press; 2015. p. 35–53.

23. Murray M. Does poor quality schooling and/or teaching hurt black South African students enrolling for a degree at the University of Kwa-Zulu Natal? PLoS ONE. 2016;11(4), e0153091, 11 pages. https://doi.org/10.1371/journal.pone.0153091

24. Zewotir Z, North D, Murray M. The time to degree or dropout amongst full-time master's students at University of KwaZulu-Natal. S Afr J Sci. 2015;111(9–10), Art. #2014-0298, 6 pages. http://dx.doi.org/10.17159/sajs.2015/20140298

25. Zewotir T, North D, Murray M. Student success in entry level modules at the University of KwaZulu-Natal. S Afr J High Educ. 2011;25(6):1233–1244.

26. Murray M. Determining the efficacy of bridging programmes in the Faculty of Science at the University of Kwa-Zulu Natal. S Afr Stat J. 2015;49:241–257.

Deflating the shale gas potential of South Africa's Main Karoo basin

AUTHORS:
Michiel O. de Kock[1]
Nicolas J. Beukes[1]
Elijah O. Adeniyi[1]
Doug Cole[2]
Annette E. Götz[3]
Claire Geel[4]
Frantz-Gerard Ossa[1]

AFFILIATIONS:
[1]DST-NRF CIMERA, Department of Geology, University of Johannesburg, Johannesburg, South Africa
[2]Council for Geoscience, Pretoria, South Africa
[3]School of Earth and Environmental Sciences, University of Portsmouth, Portsmouth, United Kingdom
[4]Department of Geological Sciences, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Michiel de Kock

EMAIL:
mdekock@uj.ac.za

DATES:
Received: 08 Nov. 2016
Revised: 10 Mar. 2017
Accepted: 10 June 2017

KEYWORDS:
Ecca Group; Whitehill Formation; hydrocarbon; thermal maturity; energy resource

HOW TO CITE:
De Kock MO, Beukes NJ, Adeniyi EO, Cole D, Götz AE, Geel C, et al. Deflating the shale gas potential of South Africa's Main Karoo basin. S Afr J Sci. 2017;113(9/10), Art. #2016-0331, 12 pages. http://dx.doi.org/10.17159/sajs.2017/20160331

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
CIMERA-KARIN

The Main Karoo basin has been identified as a potential source of shale gas (i.e. natural gas that can be extracted via the process of hydraulic stimulation or 'fracking'). Current resource estimates of 0.4–11x10^9 m^3 (13–390 Tcf) are speculatively based on carbonaceous shale thickness, area, depth, thermal maturity and, most of all, the total organic carbon content of specifically the Ecca Group's Whitehill Formation with a thickness of more than 30 m. These estimates were made without any measurements on the actual available gas content of the shale. Such measurements were recently conducted on samples from two boreholes and are reported here. These measurements indicate that there is little to no desorbed and residual gas, despite high total organic carbon values. In addition, vitrinite reflectance and illite crystallinity of unweathered shale material reveal the Ecca Group to be metamorphosed and overmature. Organic carbon in the shale is largely unbound to hydrogen, and little hydrocarbon generation potential remains. These findings led to the conclusion that the lowest of the existing resource estimates, namely 0.4x10^9 m^3 (13 Tcf), may be the most realistic. However, such low estimates still represent a large resource with developmental potential for the South African petroleum industry. To be economically viable, the resource would be required to be confined to a small, well-delineated 'sweet spot' area in the vast southern area of the basin. It is acknowledged that the drill cores we investigated fall outside of currently identified sweet spots and these areas should be targets for further scientific drilling projects.

Significance:

- This is the first report of direct measurements of the actual gas contents of southern Karoo basin shales.
- The findings reveal carbon content of shales to be dominated by overmature organic matter.
- The results demonstrate a much reduced potential shale gas resource presented by the Whitehill Formation.

Introduction

The potential shale gas resource of the Karoo Supergroup (Figure 1), and specifically the ~30-m thick Whitehill Formation of the Ecca Group, remains highly speculative.[1-7] An original ~18x10^9 m^3 or 485 trillion cubic feet resource estimate[8] – which would make the Karoo basin the fourth largest resource in the world – is certainly grossly inflated. Trillion cubic feet or Tcf is the unit in which widely published resource estimates are quoted and are provided throughout this contribution in brackets wherever resource estimates are listed in SI units. The United States Energy Information Administration downgraded this estimate to place the Karoo basin as the sixth largest global resource at 11x10^9 m^3 (390 Tcf), of which the Whitehill Formation contributed ~6x10^9 m^3 (211 Tcf).[3] Conservative estimates are much smaller. Preliminary scenarios of 0.9–8x10^9 m^3 (32–287 Tcf) were calculated as alternatives to the US estimate.[1] Subsequent work has resulted in best estimates closer to the smaller conservative value cited above. Deterministic gas estimates of 1–1.2x10^9 m^3 (36–42 Tcf) were calculated for the Whitehill Formation.[4] Comparable to this amount is the probabilistic estimate of 1.4x10^9 m^3 (49 Tcf), but with a large uncertainty interval of 0.4–4.9x10^9 m^3 (14–172 Tcf).[5] A speculated technically recoverable shale gas resource of 0.37x10^9 m^3 (13 Tcf) for the Whitehill Formation and 0.54–0.65x10^9 m^3 (19–23 Tcf) recoverable free gas represent the lower end of estimates.[2,6]

The Karoo Supergroup was deposited some 300 to 183 million years ago on the ancient continent Gondwanaland, but is now best represented by a large erosional remnant in southern Africa referred to as the Main Karoo basin (Figure 1).[9,10] Sedimentation in the basin was terminated during Gondwanaland breakup with the emplacement of the Karoo large igneous province (KLIP), which includes an extensive network of dolerite sills and dykes.[11,12] Along the basin's southern margin the Karoo succession attains a maximum composite thickness of 12 km.[10] Here the basin is bound by a narrow zone of deformation known as the Cape Fold Belt (CFB).[13] KLIP intrusions and deformation associated with the CFB distinguishes the Main Karoo basin from other well-known shale gas basins in the world.

Drilling by the Southern Oil Exploration Corporation (SOEKOR) failed to prove the existence of economic conventional hydrocarbon (particularly oil) reservoirs in the southern Main Karoo basin, but with the advent of unconventional gas plays, the basin again received attention.[1-6,8,14] However, current resource estimates may not sufficiently account for thermal degassing and possible gas escape during KLIP emplacement and development of the CFB.[15] Current estimates either include speculative risk factors to account for these effects, or are deterministic for 'sweet spot' areas where these effects are minimised. Quantitatively, however, the actual effect of KLIP intrusions and the CFB is unknown. Within the spatial limits of the current study, both the effects of KLIP intrusions and thermal tectonism of the CFB are illustrated by various maturity indices.

Unfortunately, much of the carbonaceous shales intersected by the SOEKOR cores are deteriorated and unsuitable for evaluating reservoir and source potentials. Recent studies of unweathered shale material have focused on the geothermal history and petro-physical characteristics of shale units at specific points within the basin[6,15-18], but direct measurements of the actual available gas content of the shale units are lacking.

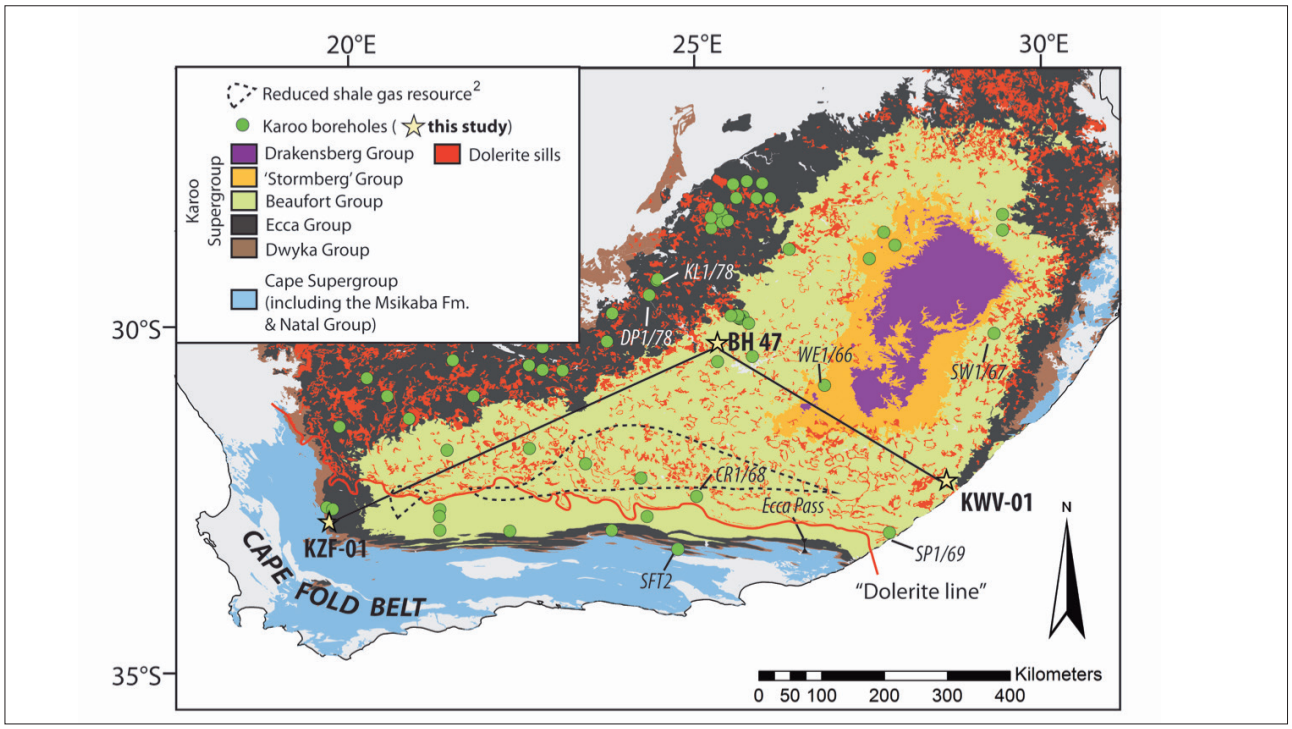


Figure 1: Simplified geological map of the Main Karoo basin showing the location of the three drill cores studied and other sites mentioned in the text.

The Karoo Research Initiative (KARIN) under the DST-NRF Centre of Excellence for Integrated Mineral and Energy Resource Analysis (CIMERA) hosted by the University of Johannesburg and co-hosted by the University of the Witwatersrand drilled two boreholes to assist in this endeavour (Figure 1; KZF-01 in the Tankwa Karoo and KWV-01 near Willowvale in the Eastern Cape Province). A borehole drilled by Gold Fields Ltd near Philippolis in the Free State Province to explore the basement rocks of the Karoo succession provides an intersection from the central part of the basin (Figure 1; BH 47).

Geological setting

Borehole KZF-01 intersects 657 m of the Ecca Group in the southwestern part of the basin near the CFB, but south of the so-called 'dolerite line' — a boundary south of which no dolerite sills are present (Figures 1 and 2). The Whitehill Formation, with a thickness of 19.5 m, was intersected at 420.46 m below the surface, but unexpected structural duplication by low-angle thrust faults or brittle deformation features resulted in further intersections between 443.30 m and 479.55 m (36 m thickness) and between 489.15 m and 498.45 m (9 m thickness). It is difficult to determine the true thickness of the Whitehill Formation given the brecciated nature of contacts, but the first 19.5 m intersection is likely to represent the true thickness as it appears that it is the Prince Albert Formation, rather than the Collingham Formation, that is tectonically displaced with the Whitehill Formation (Figure 2). The shale reservoir is too shallow to be of commercial interest as the depth of shale should be more than 1500 m for safe hydraulic fracturing. Although there is an increased risk of gas escape at depths of less than 1000 m, it should be noted that even at shallow depth any residual gas will remain in place, as will desorbed and free gas, if a suitable caprock is in place. In proximity to the CFB, the Collingham Formation overlying the Whitehill Formation has been shown to have the properties of a suitable caprock as a result of the large proportion of clay minerals, a low total organic carbon content (TOC), the fine-grained nature of lithologies, a low porosity, a lack of permeability, a moderate fracturability, average density values, and the laminate nature of the formation.[17] Near the surface (<200 m) at Loeriesfontein, in the southwestern part of the basin, gas has been escaping from the Whitehill Formation for the past 30 years after a dolerite sill above the formation was breached by a borehole.[19] In the northern part of the basin, near Evander, conventional free gas was discovered in sandstone of the Vryheid Formation at depths of less than 200 m, and only escaped because of extensive drilling.[20] There is thus very little reason to believe that borehole KZF-01 would not give a realistic reflection of the available gas in the region, despite the Whitehill Formation being at a depth of 420.46 m.

Borehole KWV-01 was drilled in an area known to contain abundant dolerite sills, but little information was available on the nature of the stratigraphic succession. The nearest reference sections are more than 100 km away, represented by SOEKOR cores SP1/69, WE1/66, and SW1/67, and outcrops along the Ecca Pass near Grahamstown (Figure 1).[21] The borehole was drilled to a depth of 2353 m commencing within the Beaufort Group, intersecting the entire Ecca Group and ending within the Dwyka Group (Figure 2). The Whitehill Formation, with its ubiquitous black carbonaceous shale, is dramatically thinner (13 m) than the average of ~39 m in other parts of the southern Karoo basin[2], and is intruded by a 19-m-thick dolerite sill (Figure 2). The low thickness of the Whitehill Formation in borehole KWV-01 renders it commercially unviable as a shale gas reservoir, and complements a predicted pinch-out of the formation some 65 km to the northeast near Coffee Bay. What was unexpected is the thick sand-dominated Ripon Formation representing much of the Ecca Group with a well-developed interstratified dark grey shale known as the Wonderfontein Member (Figure 2).

The third borehole (BH 47) from which samples were investigated intersected the Ecca and Dwyka groups down into the basement. Thick dolerite sills are characteristic features. The Whitehill Formation could not positively be identified because it is in an area immediately north of the pinch-out. However, samples near the base of the Ecca Group and above the Dwyka Group appear highly carbonaceous and thus this unit is tentatively suggested as being a lateral facies equivalent of the Whitehill Formation (Figure 2). This core is important because it (1) contains several thick dolerite sills and (2) is distal to the effects of the CFB.

Although both new KARIN boreholes were located near the present day erosional margins of the Main Karoo basin, it is important to realise that the Whitehill Formation cannot necessarily be seen as being proximal. In fact, palynofacies analyses indicate marine conditions during the deposition of the Whitehill black shales in the southern Karoo basin.

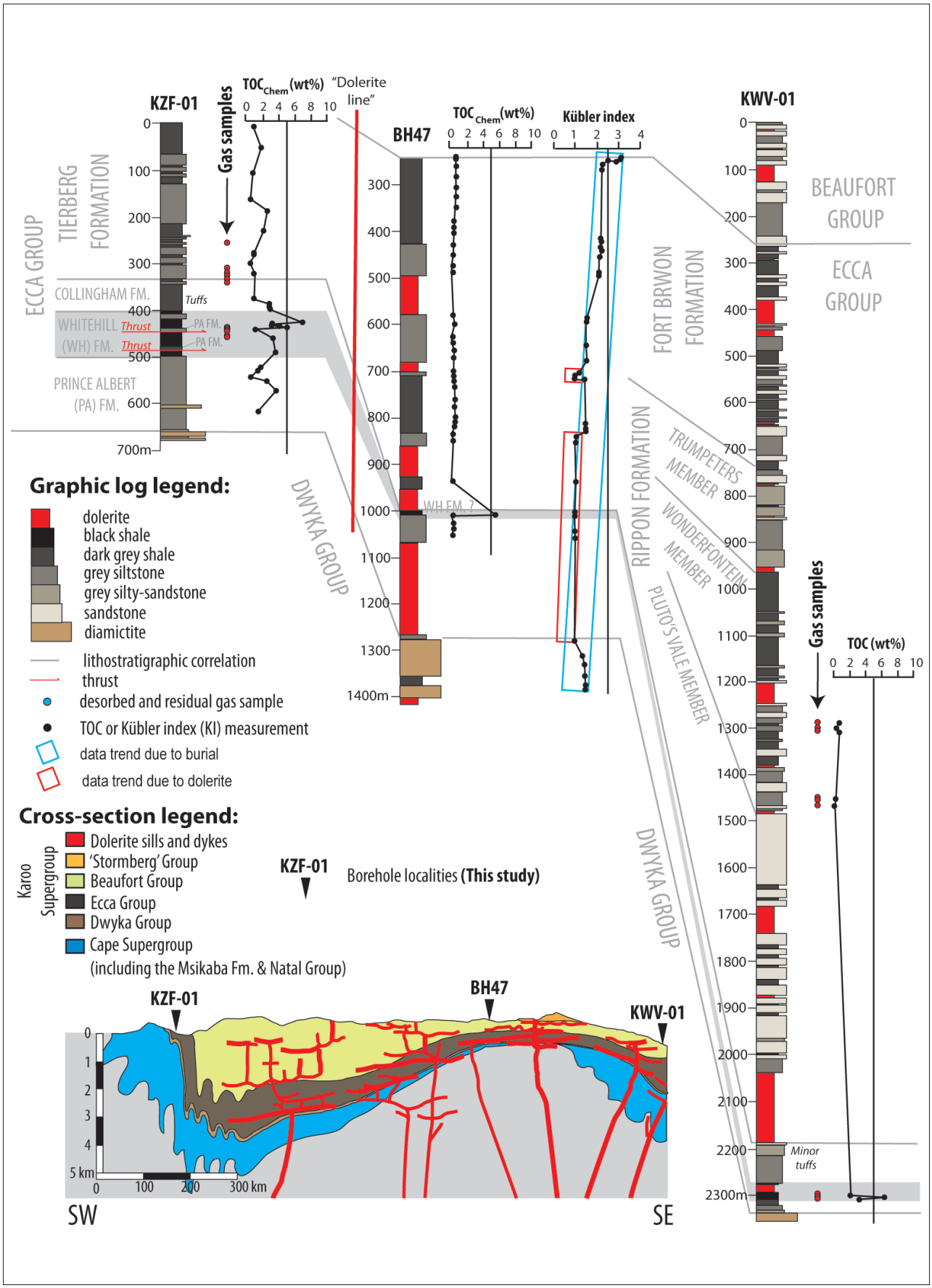


Figure 2: Simplified lithostratigraphic logs of boreholes KZF-01, BH 47 and KWV-01 showing the stratigraphic distribution of desorbed and residual gas samples, total organic carbon (TOC_{Chem}) content, and Kübler index values. Also shown is a schematic cross section (SW–SE) of the southern Karoo basin between the three cores studied.

In the southwestern part of the basin (KZF-01), palynofacies data point to a distal basinal setting with moderate marine phytoplankton percentages (i.e. acritarchs and prasinophytes), good amorphous organic matter preservation, low terrestrial input, and low spores:bisaccates ratios.[22] In the southeastern part of the basin (KWV-01), palynofacies analysis suggests a stratified deep basin setting with low marine phytoplankton percentages (i.e. prasinophytes), good amorphous organic matter preservation, high terrestrial input, and moderate spores:bisaccates ratios.[22] In contrast, a marginal marine, restricted setting was detected in the northern part of the basin (SOEKOR borehole DP 1/78) as documented by low marine phytoplankton percentages (i.e. leiospheres and prasinophytes), low amorphous organic matter preservation, high terrestrial input, and moderate spores:bisaccates ratios.[22]

Methods

In both boreholes KZF-01 and KWV-01, carbon-rich shale of the Whitehill Formation, together with a few other carbonaceous shale beds in the Ecca Group, were monitored for desorbed gas volume at the drill sites, and later analysed for desorbed gas composition and residual gas volume and composition (Figure 2). Organic carbon was characterised by measurements of TOC content and Rock-Eval pyrolysis. In borehole BH 47, the Prince Albert, Whitehill and Collingham formations could not be distinguished (Figure 2). Here, the Ecca Group is dominated by dark blue-grey shale. TOC content and Rock-Eval pyrolysis values were determined for shale samples in all three cores using the Kübler index, and vitrinite reflectance analyses were performed on samples from core BH 47 near Philippolis.

Desorbed and residual gas content and composition

Gas is generated during the maturation of organic matter in shale, and the majority of this gas is typically sorbed or attached to the surface of clay and mud particles. Upon a reduction in pressure, such as that experienced during drilling, some of the gas will desorb, which can be monitored over time at the drill site. We sampled prominent black carbonaceous shale units intersected for desorbed gas measurements on site. There were no apparent gas kicks or blow-outs detected at either of the drilling sites at any stage. Any remaining gas is residual, and it is only released during complete fracturing of the host shale by milling in a vacuum-sealed vesicle. Desorbed and residual gas content and composition of carbonaceous shales in KZF-01 and KWV-01 were monitored by Geokrak (Poland) and by Latona Consulting (South Africa).

For desorption analyses of KZF-01, 20 core samples each of about 300 mm in length were selected from carbonaceous units and transferred to leak-tight stainless steel canisters in a Geokrak field laboratory immediately after sampling (Figure 3a). The time elapsed between starting a drill run, core retrieval, and eventual sample selection was carefully monitored to account for any lost gas. Accounting for lost gas can have a large effect, but in this case it did not alter the results significantly. Air was removed from the canisters by displacement with helium from a pressurised cylinder. The canisters were closed tight with an expansive plug, weighed and placed in a thermostatic heater. The desorbed gas volume, released as samples were allowed to equilibrate to ambient temperatures in a thermostatic heater, was measured with a volumeter at set time intervals. Initially the readings were made at regular, short intervals. As the gas volume diminished, the interval between readings was lengthened. Desorption of cores was terminated when a single reading of gas volume measured in a 24-h cycle was smaller than 5 cm^3, or if the amount of desorbed gas released by a core sample in 7 days was less than 1% of the total gas desorbed from the sample. The amount of gas released from core samples was expressed in volume unit per mass unit.

Desorption analyses of KWV-01 was comparably accomplished by Latona Consulting at the drill site. They used leak-tight PVC canisters without displacement of air with helium and without the use of a thermostatic heater. Lost gas, or gas released before samples were sealed, was calculated graphically. Plotting the cumulative desorbed gas in millilitres against the square root of time produced a straight line for about 10 h after coring, and the straight line was projected backwards before the time when the canister was sealed to estimate the lost gas.

One sample of desorbed gas was selected on site from KWV-01 via a plastic pipette for analysis of its content by gas chromatography at the South African Nuclear Energy Corporation (NECSA) in Pretoria, South Africa.

For the residual gas measurements, pieces from each core sample (KZF-01 and KWV-01) were collected when desorption was finished, and milled in a leak-tight stainless steel vessel (Figure 3b) at the respective laboratories of Geokrak and Latona Consulting in Poland and Johannesburg. Measurements of residual gas were made with a volumeter after specified time intervals of milling. The standard milling time was extended from 60 min to 120 min if no gas was released.

Pipettes of residual gas from KZF-01 were collected during residual gas analyses and were subsequently analysed for contents by gas chromatography at NECSA in Pretoria, South Africa, and at the Oil and Gas Institute in Cracow, Poland.

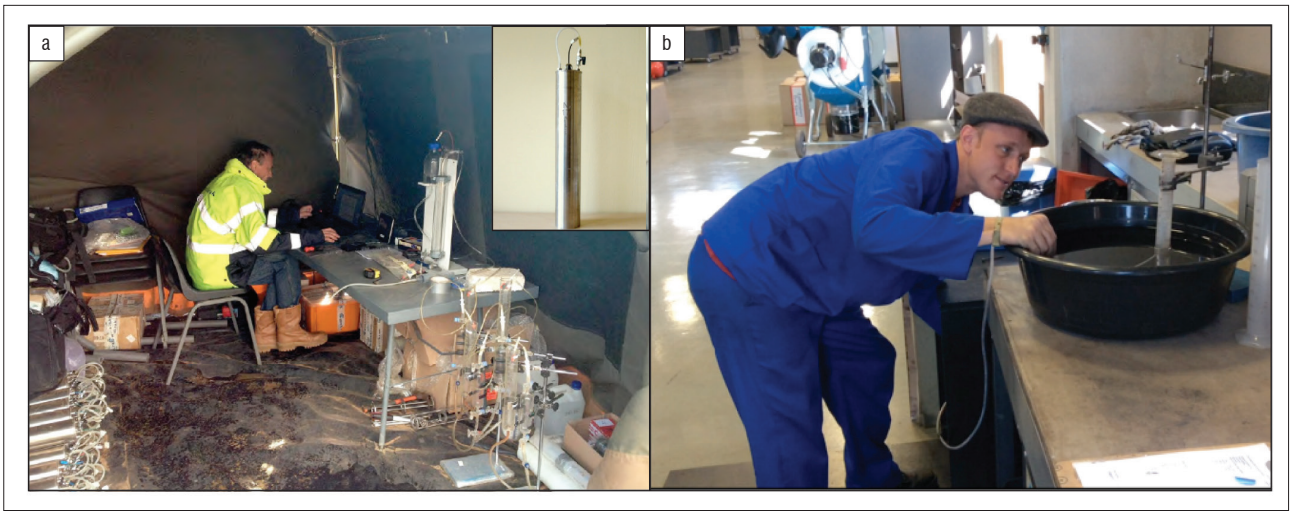


Figure 3: (a) Geokrak's desorption field laboratory at the KZF-01 drill site. The inset shows a leak-tight stainless steel desorption canister. (b) Measurement of residual gas content of a milled shale sample at Latona Consulting's Johannesburg laboratory.

Total organic carbon content 'chemical' method

For the total organic carbon content 'chemical' method (TOC_{Chem}), carbonaceous shale samples were selected from cores of boreholes BH 47 (38 samples) and KZF-01 (26 samples). Samples were cut and milled and 5 g of rock powder per sample was selected for analysis. Samples from BH 47 were analysed at the Institute for Geology and Palaeontology of the University of Münster in Germany. The samples from KZF-01 were analysed at the Department of Geology, University of Maryland in the USA. The TOC content was determined via sealed tube combustion.[23] Between 10 mg and 500 mg of rock powder was decarbonated in a quartz tube with HCl (25%), washed to neutrality and dried at 40 °C. Subsequently, ca 1.5 g of CuO was added and the quartz tubes were sealed under vacuum. CO_2 was liberated from the sample powder via combustion at 850 °C for 3 h, cryogenically purified, quantified and packed in a 6-mm break-seal tube. Analytical performance was monitored using several international (USGS 24, IAEA 40) and in-house laboratory (coal) standards.

Vitrinite reflectance

Seven samples were selected from BH 47 of carbonaceous shale units both proximal and distal to the dolerite intrusions. Samples were prepared according to the ASTM D7708–14 standard test method for microscopic determination of the reflectance of vitrinite dispersed in sedimentary rocks.[24] Whole-rock samples were mounted in 30-mm moulds with epoxy resin and allowed to cure overnight. Individual mounts were polished to produce a smooth plane surface using a Struers Tegramin polisher. The random reflectance measurement procedures[24] were followed using a Zeiss Axio Imager M2, retrofitted with a Hilgers Fossil Diskus system. Both petrographic identification and vitrinite reflectance readings were determined under non-polarised light. Mean random vitrinite reflectance (R_oVmr) was measured in percentages of the intensity of reflected light illuminated on a polished plane surface of the rock sample covered with immersion oil by a calibrated microscope or photometer with an x50 oil objective. An average of 32 to 98 vitrinite measurements were taken per sample, depending on the availability of organic matter, and the mean values were determined.

Kübler index

The Kübler index (KI)[25] was determined using X-ray diffraction analysis at the University of Johannesburg's SPECTRUM. X-ray diffraction analyses were performed using the Panalytical X'Pert Pro X-ray diffractometer with an X'Celerator detector, the CuKα radiation operated at 40 kV and 40 mA. KI was determined for oriented clay particles ($<2\ \mu$m) separated from six samples. Air-dried oriented clay separates ($<2\ \mu$m particles) were prepared by placing mildly crushed sample material in lidded bottles that were half-filled with osmosis water before being placed in an ultrasonic bath for over 3 h for separating clay particles from the detrital minerals (e.g. quartz and feldspars). The bottles were placed in a fume box for a minimum of 8 h to allow the solutions to attain room temperature or ~20 °C. The solution was then shaken and left for 2.5 h to allow a suspension of $<2\ \mu$m particles from the solution according to Stoke's Law.[26] The water with suspended clay particles was pipetted into a clean beaker and placed in an oven at 40 °C to dry out the water and collect the fine clay-rich powders.

KI was also determined from 37 bulk rock samples. KI is calculated as the width at half-height of an illite peak at 10 Å. Results obtained from clay-rich separates are consistent with those from bulk rock analysis. Therefore, KI values of bulk rocks were used to complete the KI trend across the borehole.

Rock-Eval pyrolysis

Shale samples were evaluated using a Rock-Eval 6 pyrolyser at the Department of Earth Sciences of the Indian Institute of Technology, India. Powdered sample material was pyrolysed in an inert atmosphere and the residual carbon was subsequently burnt in an oxidation oven. The amount of hydrocarbons released (S_1 and S_2) during the pyrolysis between 300 °C and 650 °C, later increased to 750 °C at a rate of 25 °C/min, were detected with a flame ionisation detector. Free hydrocarbons are designated as S_1 and hydrocarbons generated with further thermal cracking are designated as S_2. The temperature at which hydrocarbon yield is maximised is termed T_{max}. The gases released during the pyrolysis [CO and CO_2 (S_3)] were detected with an online infrared detector continuously throughout the process.[27] Any remaining carbon after pyrolysis is residual (S_4). The TOC content from pyrolysis (TOC_{Pyro}) is not directly measured, but can be calculated as a weight percentage using Equation 1:

$$TOC_{Pyro} = [0.082(S_1 + S_2) + S_4]/10, \quad \text{Equation 1}$$

where 0.082 is a constant representing the average amount of carbon from thermally extracted and pyrolysed hydrocarbons.[28]

Several indices can be calculated to evaluate the geochemistry of the organic matter as well as its thermal maturity.[27] The hydrogen index or HI, determined by Equation 2, provides a measure of the relative amount of organic matter still capable of producing petroleum, sometimes referred to as 'live' organic matter. The production index or PI is calculated using Equation 3 and provides an estimate of the extent to which oil generation has taken place. The oxygen index or OI is defined by Equation 4 and provides a measure of the amount of organic bound oxygen in the sample.

$$HI = S_2/TOC_{Pyro} \times 100 \quad \text{Equation 2}$$

$$PI = S_1/(S_1+S_2) \quad \text{Equation 3}$$

$$OI = S_3/ TOC_{Pyro} \times 100 \quad \text{Equation 4}$$

Results

Desorbed gas contents of samples were very low (Table 1). The largest volume (0.22 m^3/t) was obtained from the Wonderfontein Member in KWV-01, but was only a small initial desorbed volume. Desorbed gas was essentially carbon dioxide with very little methane at a concentration of 4.8 ppm. Samples yielded little or, as was the case for KWV-01, no residual gas. KZF-01 yielded inconsistent residual gas volumes (0.00–0.74 m^3/t; Table 1). The Whitehill Formation did not contain elevated gas content. Residual gas was mostly methane (61–99%), with variable concentrations of nitrogen and carbon dioxide (Table 2).

TOC_{Chem} of samples ranges between 0.01 wt% and 6.83 wt% (Figure 2; Table 3). Content is generally low for shale samples of the Tierberg (0.44–2.54 wt%) and Collingham formations (0.91–2.87 wt%) in KZF-01 (Table 3) and higher for the Whitehill Formation (1.19–6.83 wt%). The Prince Albert Formation has a very variable, but overall low TOC_{Chem} (0.47–3.64 wt%). TOC_{Chem} is very low in BH 47 (0.04–0.42 wt.%), but one sample considered correlative with the Whitehill Formation at a depth of 1011.25 m yielded 5.59 wt%. This concentration is comparable to that of the Whitehill and Prince Albert formations in KZF-01. The Whitehill Formation's average TOC_{Chem} content in our boreholes (i.e. 3.77 wt% in KZF-01 based on eight samples and 5.59 wt% in one sample from BH 47) is generally above the 2 wt% qualifying value employed in original shale gas resource estimation, but lower than the 6 wt% average on which resource estimates were based.[1,8]

Vitrinite reflectance measurements (Figure 4) of BH 47 Ecca Group shale samples that lie further away from dolerite sills display unexpectedly higher values (3.71–3.91%), compared to those closer to dolerite sills (1.17–1.77%). Organic matter fragments are rare and generally very small. Samples far away from dolerite intrusions appear to have no structure or orientation, have very fine-grained and shattered organic matter amongst coarser quartz grains and framboidal pyrite (Figure 5a and 5b), while samples closer to intrusions have an apparent orientation of organic matter, which is layered and networked around quartz particles with pyrite inclusions (Figure 5c and 5d). The organic matter is generally highly matured and appears as solid bitumen networks, and is more likely to be inertinite than vitrinite. Reflectance values are thus better referred to as total reflectance rather than vitrinite reflectance.

Table 1: Gas content of KZF-01 and KWV-01 core samples

Borehole	Formation or Member	Sample number	Core interval (in m)		Desorbed gas	Residual gas	Total gas
			from	to	(in m^3/t)		
KZF-01	Tierberg	BIZ-84/01/D	262.08	262.38	0.01	0.12	0.13
		BIZ-84/02/D	312.26	312.56	0.01	n.a.	n.a
		BIZ-84/03/D	319.38	319.77	0.01	n.a.	n.a
		BIZ-84/04/D	323.05	323.45	0.01	0.41	0.45
		BIZ-84/05/D	329.10	329.40	0.01	n.a.	n.a
	Collingham	BIZ-84/06/D	340.83	341.13	0.00	0.27	0.27
	Whitehill	BIZ-84/07/D	422.10	422.34	0.00	0.11	0.11
		BIZ-84/08/D	423.32	423.62	0.00	0.24	0.24
		BIZ-84/09/D	425.10	425.40	0.00	0.18	0.18
		BIZ-84/10/D	426.24	426.56	0.00	n.a.	n.a
		BIZ-84/11/D	428.10	428.38	0.00	n.a.	n.a
		BIZ-84/12/D	429.10	429.40	0.00	n.a.	n.a
		BIZ-84/13/D	431.10	431.39	0.01	0.22	0.23
		BIZ-84/14/D	432.29	432.57	0.00	0.17	0.17
		BIZ-84/15/D	434.04	434.34	0.01	n.a.	n.a.
		BIZ-84/16/D	435.55	435.85	0.00	n.a.	n.a.
		BIZ-84/17/D	437.08	437.38	0.00	0.00	0.00
		BIZ-84/18/D	438.54	438.82	0.01	n.a.	n.a.
		BIZ-84/19/D	447.80	448.17	0.01	0.74	0.75
		BIZ-84/20/D	449.35	449.64	0.00	0.56	0.56
KWV-01	*Wonderfontein*	LT01	1291.27	1292.27	0.20	0.00	0.20
		LT02	1303.27	1304.27	0.00	0.00	0.00
		LT03	1309.27	1310.27	0.03	0.00	0.03
	Pluto's Vale	LT04	1450.27	1451.27	0.02	0.00	0.02
		LT05	1453.27	1454.27	0.01	0.00	0.01
		LT06	1465.27	1466.27	0.05	0.00	0.05
	Whitehill	LT07	2295.02	2295.52	0.01	0.00	0.01
		LT08	2299.39	2299.59	0.00	0.00	0.00
		LT09	2305.39	2305.89	0.00	0.00	0.00

n.a., not analysed

Table 2: Residual gas composition in KZF-01

Formation	Sample number	Air free gas components (%)		
		CH_4	N_2	CO_2
Tierberg	BIZ-84/04/D	99.10	0.00	0.90
Collingham	BIZ-84/06/D	99.72	0.00	0.28
Whitehill	BIZ-84/08/D/R	68.89	7.59	22.70
	BIZ-84/13/D	61.38	38.46	0.15
	BIZ-84/14/D/R	68.78	7.90	22.17
	BIZ-84/19/D	83.91	0.00	26.09
	BIZ-84/20/D/R	85.83	3.45	10.24

Table 3: Total organic carbon (TOC_{chem}) content of shale samples

Borehole	Formation	Sample number†	TOC_{Chem} (in wt%)
KZF-01	Tierberg	KZF-14.36	0.85
		KZF-55.74	1.77
		KZF-113.79	0.72
		KZF-164.41	0.44
		KZF-189.38	2.54
		KZF-231.7	2.08
		KZF-275.95	0.96
		KZF-279.81	1.12
		KZF-299.1	0.36
		KZF-323.59	1.01
	Collingham	KZF-376.82	0.91
		KZF-385.58	2.78
		KZF-398.10	2.87
	Whitehill	KZF-424.5	6.83
		KZF-428.79	3.38
		KZF-431.36	3.17
		KZF-431.65	3.79
		KZF-434.34	5.02
		KZF-438.82	1.19
		KZF-458.1	3.23
		KZF-488.1	3.55
	Prince Albert	KZF-518.1	1.93
		KZF-540.43	0.47
		KZF-549.08	2.47
		KZF-568.63	3.64
		KZF-611.76	1.34
BH 47	Undifferentiated	BH47-242.80	0.35
		BH47-244.50	0.33
		BH47-247.40	0.37
		BH47-261.37	0.39
		BH47-283.15	0.39
		BH47-308.15	0.42
		BH47-328.00	0.37
		BH47-350.82	0.32
		BH47-381.97	0.20
		BH47-392.40	0.19
		BH47-408.50	0.15
		BH47-434.50	0.07
		BH47-455.50	0.08
		BH47-478.50	0.05
		BH47-495.00	0.05
		BH47-586.00	0.07
		BH47-605.00	0.23
		BH47-632.00	0.06
		BH47-644.27	0.05
		BH47-660.60	0.19
		BH47-678.00	0.14
		BH47-706.00	0.07
		BH47-716.09	0.13
		BH47-726.02	0.18
		BH47-741.12	0.24
		BH47-766.64	0.28
		BH47-782.56	0.26
		BH47-804.51	0.31
		BH47-814.82	0.25
		BH47-824.83	0.25
		BH47-842.00	0.05
		BH47-854.30	0.08
		BH47-939.00	0.04
	Whitehill?	BH47-1011.25	5.59
	Undifferentiated	BH47-1018.00	0.06
		BH47-1033.00	0.09
		BH47-1047.00	0.13
		BH47-1060.50	0.05

†*Sample numbers correspond to depth in metres*

Published vitrinite reflectance data from the Main Karoo basin are limited, but suggest general increasing maturity (R_o = 1.0% to 4.3%) for the Whitehill Formation from the north to the south of the basin for samples unaffected by dolerite sills – a trend that reflects the tectonic front of the CFB.[6,14,29] Samples affected by dolerite sills exhibit a higher reflectance of up to 8.8%.[29] The total reflectance values obtained here for Ecca Group shales from BH 47 fit the expectation, but the very low total reflectance values from Ecca Group shale in proximity to dolerite sills is unexpected. The fine-grained nature of organic matter and lack of clearly identifiable vitrinite in samples close to dolerites place a caution on these measurements.

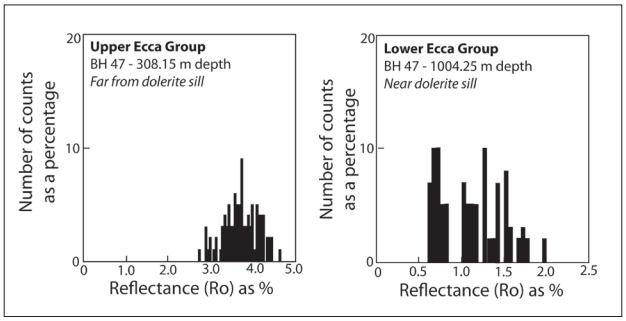
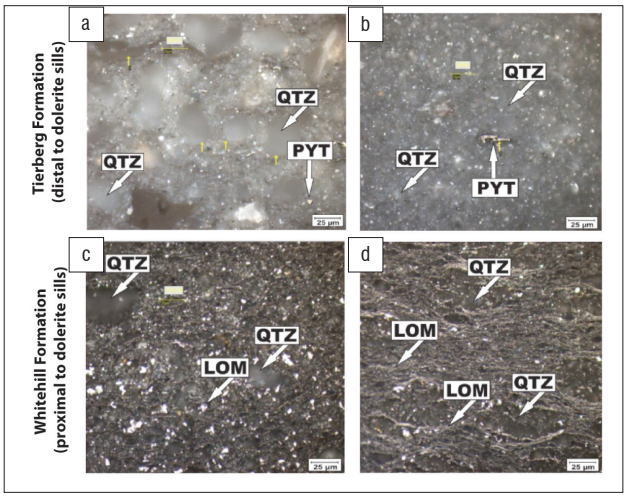


Figure 4: Examples of vitrinite reflectance data plotted as histograms for carbonaceous shale samples from borehole BH 47 distal from dolerite sills and in proximity to dolerite sills.



QTZ, quartz; PTY, pyrite; LOM, layered organic matter.

Figure 5: Petrographic images of carbonaceous shale samples from BH 47. (a) Upper Ecca Group shale (Tierberg Formation?) at 308.15 m depth distal from dolerite sills. (b) Upper Ecca Group shale (Tierberg Formation?) at 766.64 m depth distal from dolerite sills. (c) Whitehill Formation equivalent at 1004.63 m depth in proximity to a dolerite sill. (d) Whitehill Formation equivalent at 1011.25 m depth in proximity to a dolerite sill.

A progressive decrease of the KI values represents a gradual increase in non-expandable illite layers and the disappearance of the expandable smectite layers in the smectite-illite mixed layers as depth increases.[30] Within the Main Karoo basin, KI values reveal a north–south increasing effect of burial maturity and range from >5 in the north to >3 in the south for shales from outcrops.[14,18] In BH 47, KI values range from 3.15 near the surface to 1.50 at a depth of 1385 m (Figure 2; Table 4). Most of the samples yield values below 2.50, which marks the onset of metamorphic conditions (Figure 2). In addition, a local trend is seen with KI dropping to as low as 1.00 as contacts with dolerite sills are approached (Figure 2).

Table 4: Kübler index of shale samples from BH 47 with relative stratigraphic position of dolerite sills indicated

Borehole	Formation	Sample number†	Kübler index
BH 47	Undifferentiated	BH47-242.80	3.15
		BH47-244.50	3.10
		BH47-247.40	2.54
		BH47-251.70	2.36
		BH47-255.86	2.30
		BH47-267.26	2.25
		BH47-416.58	2.20
		BH47-422.01	2.20
		BH47-434.50	2.18
		BH47-455.50	2.15
		BH47-489.00	2.10
		BH47-495.00	2.10
		Dolerite sill	
		BH47-586.00	1.58
		BH47-595.00	1.55
		BH47-644.27	1.53
		BH47-678.00	1.51
		Dolerite sill	
		BH47-706.00	1.20
		BH47-711.59	1.00
		BH47-715.00	1.00
		BH47-716.09	1.00
		BH47-718.99	1.40
		BH47-819.32	1.47
		BH47-824.83	1.50
		BH47-830.00	1.50
		BH47-842.00	1.09
		BH47-854.30	1.05
		Dolerite sill	
		BH47-932.00	1.05
		Dolerite sill	
	Whitehill?	BH47-1004.6	1.01
		BH47-1011.25	1.01
	Undifferentiated	BH47-1047.00	1.00
		BH47-1060.50	1.00
		Dolerite sill	
		BH47-1280.70	1.00
		BH47-1313.11	1.30
		BH47-1333.00	1.35
		BH47-1357.84	1.40
		BH47-1377.94	1.48
		BH47-1385.12	1.50

†*Sample numbers correspond to depth in metres*

Pyrograms obtained during Rock-Eval pyrolysis analyses of carbonaceous shale samples reveal low amounts of free hydrocarbon (S_1) and poorly defined S_2 (hydrocarbons released by thermal cracking) peaks (Figure 6), which results in unreliable constraints of T_{max} and low thermal maturity indices such as the hydrogen and production indices (Table 5). The hydrocarbon generation potential of the organic matter or kerogen is generally poor (calculated as the sum of S_1 and S_2; Figure 7a) despite promising TOC_{Pyro} contents calculated from pyrolysis. The low hydrogen index suggests that much of the organic matter is not bound to hydrogen, and that hydrocarbon generation could have taken place in the past. Much of the organic carbon is thus 'dead' carbon. If hydrocarbon generation occurred in the basin, then it was not readily preserved as suggested by the low production index and the low volumes of residual gas. Organic matter or kerogen is of poor quality in terms of hydrocarbon generating potential according to a scheme that compares the production index with TOC_{Pyro}.[31] Poor quality kerogen is also seen elsewhere in the basin (Figure 7a).[6,29,32] Generally, kerogen is either gas-prone Type III kerogen or Type IV kerogen (Figure 7b). The former is the likely final residue of a pre-existing kerogen type that has completely matured (i.e. 'dead' organic carbon). However, kerogen in borehole DP1/78 near Hopetown in the northern part of the basin (Figure 1) displays a thermal evolution trend of a Type I kerogen (oil prone), the maturation trend of which is now within the wet and dry gas domain (Figure 7).[33] This finding further supports the overmature nature of pre-existing kerogen.

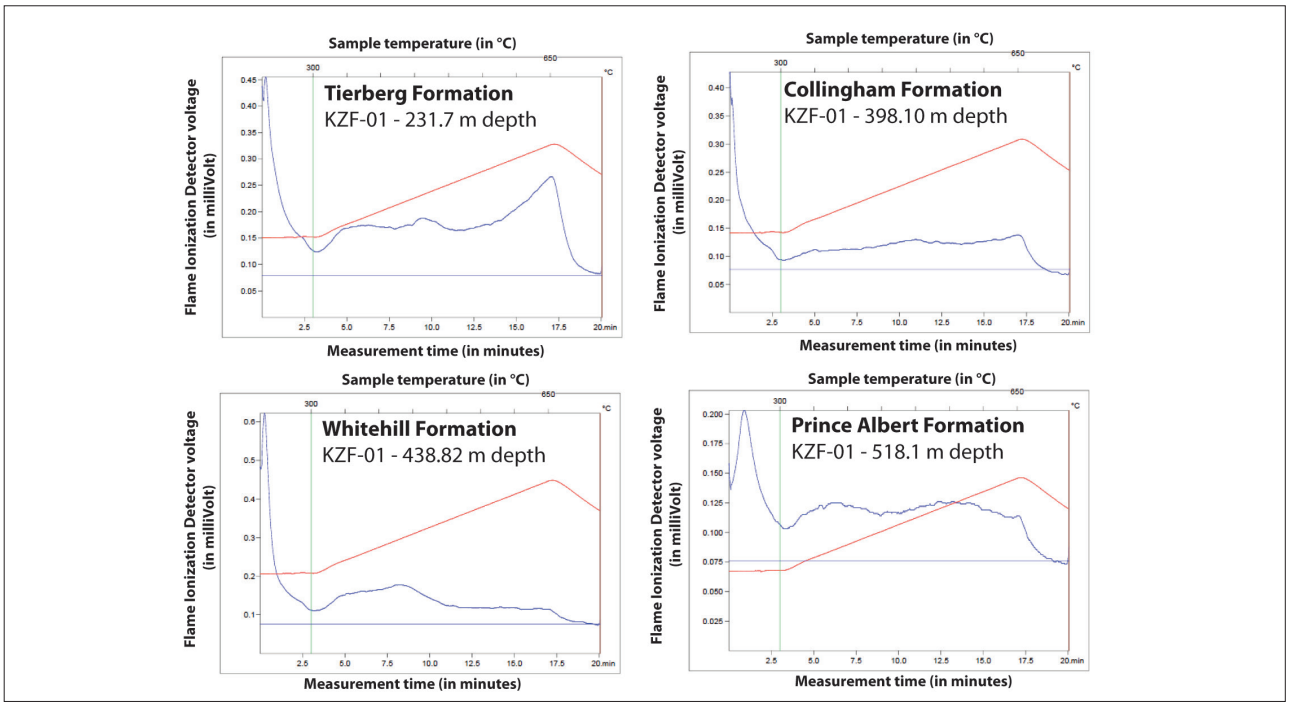


Figure 6: Selected pyrograms obtained of carbonaceous shale samples from KZF-01.

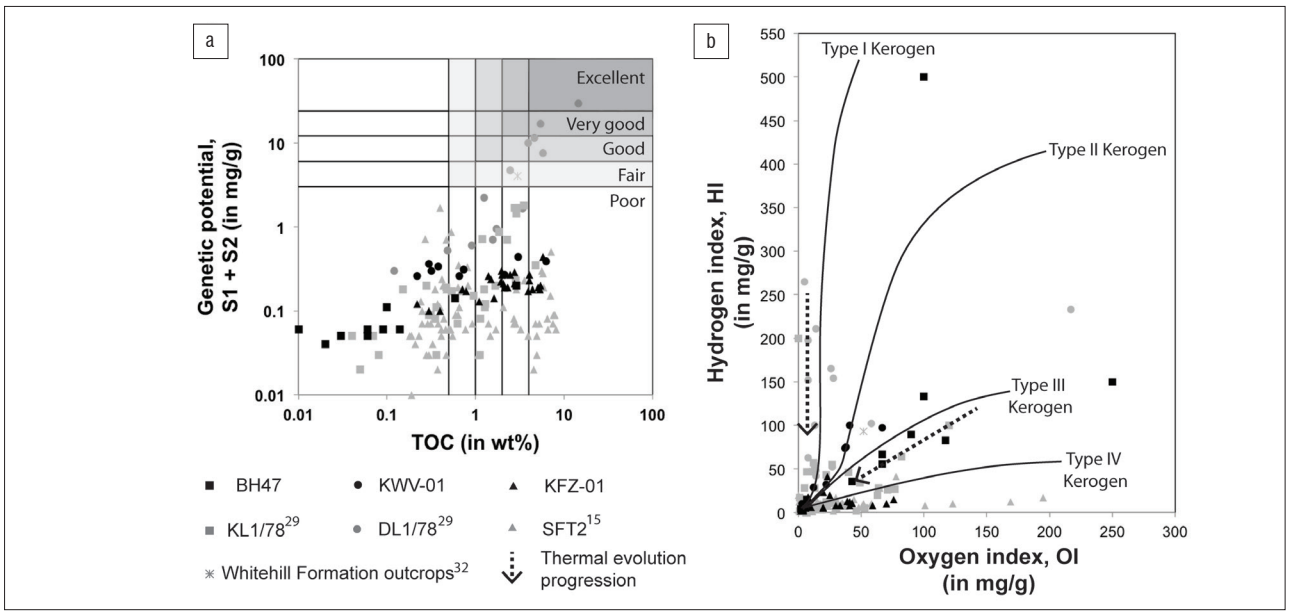


Figure 7: (a) Classification of kerogen quality in the Ecca Group carbonaceous shales.[31] (b) A modified Krevelen diagram[31] indicates the dominance of Type III and IV kerogen in Ecca Group shales of KZF-01, BH 47 and KWV-01 with reference to other studies in the Karoo basin as indicated.

Table 5: Summary of Rock-Eval pyrolysis results

Borehole	Formation or Member	Sample number†	S_1	S_2	T_{max}	S_3	TOC_{Pyro} ‡	HI	OI	PI
			(mg/g)		(°C)	(mg/g)	(wt%)			
KZF-01	Tierberg	KZF-14.36	0.06	0.18	604	0.61	1.5	12	41	0.25
		KZF-55.74	0.08	0.18	603	0.55	1.41	13	39	0.31
		KZF-113.79	0.05	0.12	607	0.59	0.78	15	76	0.29
		KZF-164.41	0.08	0.22	604	0.84	1.97	11	43	0.27
		KZF-189.38	0.08	0.19	605	1.38	1.98	10	70	0.30
		KZF-231.7	0.06	0.23	604	1.13	2.77	8	41	0.21
		KZF-275.95	0.05	0.16	605	0.62	2.07	8	30	0.24
		KZF-279.81	0.05	0.14	606	0.45	2.3	6	20	0.26
		KZF-299.1	0.07	0.14	605	0.56	2.77	5	20	0.33
		KZF-323.59	0.08	0.19	605	0.8	2.45	8	33	0.30
	Collingham	KZF-376.82	0.05	0.14	605	0.12	2.13	7	6	0.26
		KZF-385.58	0.04	0.1	604	0.16	1.63	6	10	0.29
		KZF-398.10	0.04	0.09	602	0.65	1.1	8	59	0.31
	Whitehill	KZF-424.5	0.08	0.12	388	0.2	5.46	2	4	0.40
		KZF-428.79	0.08	0.1	322	0.17	4.37	2	4	0.44
		KZF-431.36	0.05	0.13	396	0.1	5.33	2	2	0.28
		KZF-431.65	0.06	0.21	607	0.25	4.06	5	6	0.22
		KZF-434.34	0.06	0.17	392	0.17	4.06	4	4	0.26
		KZF-438.82	0.07	0.12	382	0.1	5.11	2	2	0.37
		KZF-458.1	0.04	0.13	605	0.14	3.92	3	4	0.24
		KZF-488.1	0.08	0.36	392	0.23	5.76	6	4	0.18
	Prince Albert	KZF-518.1	0.03	0.09	488	0.05	0.22	41	23	0.25
		KZF-540.43	0.02	0.08	491	0.1	0.4	20	25	0.20
		KZF-549.08	0.03	0.07	503	0.06	0.3	23	20	0.30
		KZF-568.63	0.05	0.13	344	0.06	0.71	18	8	0.28
		KZF-611.76	0.06	0.16	304	0.09	1.93	8	5	0.27
BH 47	Undifferentiated	BH47-369.51	0.02	0.09	416	0.09	0.1	90	90	0.18
		BH47-416.58	0.01	0.05	492	0.07	0.06	83	117	0.17
		BH47-711.59	0.01	0.03	336	0.05	0.02	150	250	0.25
		BH47-756.64	0.01	0.05	493	0.06	0.09	56	67	0.17
		BH47-790.25	0.01	0.05	584	0.06	0.14	36	43	0.17
		BH47-819.32	0.01	0.04	494	0.04	0.06	67	67	0.20
	Whitehill?	BH47-1011.25	0.08	0.12	300	0.07	2.87	4	2	0.40
	Undifferentiated	BH47-1274.80	0.01	0.05	491	0.01	0.01	500	100	0.17
		BH47-1360.44	0.01	0.04	492	0.03	0.03	133	100	0.20
		BH47-1377.94	0.05	0.09	449	0.04	0.59	15	7	0.36
KWV-01	*Wonderfontein*	KWV-1291.27	0.07	0.19	481	0.08	0.66	29	12	0.27
		KWV-1303.27	0.06	0.28	441	0.14	0.38	74	37	0.18
		KWV-1309.27	0.07	0.24	498	0.16	0.74	32	22	0.23
	Pluto's Vale	KWV-1450.27	0.06	0.24	427	0.12	0.32	75	38	0.20
		KWV-1453.27	0.07	0.29	473	0.2	0.3	97	67	0.19
		KWV-1465.27	0.04	0.22	437	0.09	0.22	100	41	0.15
	Whitehill	KWV-2295.02	0.06	0.21	338	0.07	2.12	10	3	0.22
		KWV-2299.39	0.12	0.27	429	0.23	6.27	4	4	0.31
		KWV-2305.39	0.16	0.28	299	0.22	3.04	9	7	0.36

†Sample numbers correspond to depth in metres. ‡TOC_{Pyro} determined during Rock-Eval pyrolysis.

Discussion

Our data indicate that heating by dolerite sills in boreholes KWV-01 and BH 47 and burial metamorphism in boreholes KZF-01 and BH 47 have resulted in the elevated thermal maturity of organic matter and the destruction of hydrocarbon potential in our samples. Negligible amounts of desorbed and residual gas volumes suggest that natural gas, if generated sometime in the geological history of the strata (rock succession), was not preserved in KZF-01 and KWV-01. Although gas generation probably did occur, much gas was likely lost via thermal degassing at KWV-01 and tectonic deformation (possibly through thrusting with increased load) at KZF-01. Contact metamorphism by dolerite sills has resulted in catastrophic and explosive degassing and alteration of shale, but no textural evidence was encountered in the boreholes of this study.[34] A high thermal gradient can lead to overmaturity of organic matter and the production of bitumen derived from hydrocarbons, which can migrate and solidify in fractures.[35] Although such vein-like deposits were not encountered in our boreholes, it was noted that organic matter in the black shale at 1011.25 m in BH 47 appears as solid bitumen networks of inertinite.

The reservoir potential of high maturity shales is not well understood, but for the Marcellus Shale of the Appalachian basin in North America it has been suggested that regional metamorphism may have destroyed the shale reservoir's porosity and self-sealing capacity.[36] A high thermal gradient related to burial, tectonism or increased mantle heat flow may have overmatured the organic matter in the Karoo shales, such that the porosity and self-sealing capacity of the Whitehill Formation was destroyed leading to the non-preservation of gas. Gas was not preserved as shale gas in any of our boreholes. It is possible that gas was generated, but this gas likely escaped to be trapped locally in rare instances. Such trapped conventional gas was most probably encountered in SOEKOR borehole CR 1/68 (Figure 1)[14], where gas was found hosted in fractured shale of the Fort Brown Formation. This gas could have escaped from the underlying Whitehill Formation. Very small gas and oil shows have been recorded from the less mature northern part of the Main Karoo basin.[14,20] However, high-volume gas shows like that of CR 1/68 have not been encountered often, suggesting that it is a rare occurrence.

It is unclear at this stage if there remain significant areas of conventional and unconventional gas retention by shale in the basin. Such areas are likely low-risk 'sweet spots'[2,4,5], and are expected to have lower thermal maturity, to be relatively free of dolerite intrusions, and to have additional suitable attributes (e.g. in respect of burial depth and thickness). It should be noted that CR 1/68 is located just south of such a sweet spot defined by low-risk segment mapping (i.e. low or no dolerite volume in the Whitehill Formation).[2,5] The apparent rare occurrence of higher volume gas shows is thus perhaps a function of the small number of boreholes that intersect such low-risk sweet spots. The most recent attempts to estimate the shale gas resource of the Karoo basin have endeavoured to identify such sweet spots by interpolating sparse legacy data.[2,4-6] These areas represent the most realistic shale gas resource in the Main Karoo basin, but their real gas potential remains to be tested. Furthermore, the confident identification of such areas ultimately requires an as yet unreached level of understanding of dolerite sill distribution and dynamics throughout the basin. Such understanding is subject to the acquisition of high-resolution geophysical data.

Another significant outcome of the KARIN research is that TOC values are variable and generally lower than the average values used during original shale gas resource estimations.[1,8] The distal setting of the Whitehill Formation in KZF-01 and KWV-01 would suggest that the determined TOC values are representative. With an average TOC content of 3.77 wt%, the Whitehill Formation in KZF-01 (arguably the most distal of all our samples) is significantly lower than the 6 wt% average used in original estimations, but perhaps more comparable to the 4.55 wt% estimate for sweet spots.[2] Much more important are Rock-Eval analyses results that indicate most of the organic matter to be fully burned or 'dead' carbon, with little hydrocarbon-generating potential remaining. The carbonaceous shales at some stage after their deposition likely yielded large volumes of hydrocarbons but very little of that appears to have been preserved at the sites investigated, other than the occurrence of sparse, fracture-filled, vein-like deposits of pseudocoal, which represent solidification of viscous bitumen derived from hydrocarbons.[35]

Conclusion

Regional heterogeneity in organic carbon content and thermal maturity, although accounted to some extent in shale gas resource estimates of the southern Main Karoo basin, may pose a greater risk towards low hydrocarbon generation and preservation. Our results indicate that carbon content is very variable, even within the same formation at a specific locality within the basin. The original resource estimates are thus likely highly inflated, although even much lower estimates are of commercial interest, given that, for example, the Mosgas project was initiated on an initial resource (reserve base) of 0.03×10^9 m^3 (1.0 Tcf).[37] We show that hydrocarbon generation and preservation is negatively affected by an elevated thermal gradient and dolerite intrusions. In addition to the localised effect of dolerite intrusions, we identify the significant risk posed by an increased thermal gradient (as a result of burial, tectonism or increased mantle heat flow) towards shale gas generation and preservation. A high thermal gradient likely resulted in overmaturity of organic matter in the Whitehill Formation and destruction of its porosity and self-sealing capability. Quantification of the real resource should be limited to 'thermal oases', in which shale gas was generated and preserved. Such areas are expected to have low or no dolerite volume in the Whitehill Formation. Very likely the most realistic resource estimates for the Karoo basin are between 0.37×10^9 m^3 and 1.4×10^9 m^3 (13–49 Tcf), with the lower estimate perhaps being the most realistic given the sparsity of data.[2,4-6] This estimation is supported by our study, in which samples distal from dolerite sills in BH 47, situated in an area far away from the CFB, are also overmature. Thus, it is expected that rocks in so-called sweet spots near Beaufort West and Sutherland (but closer to the CFB) should show the same, or even higher, maturity.

A comparison of data from the Whitehill Formation with that of the shale gas producing Barnett Shale in Texas indicates a gas resource of about 0.37×10^9 m^3 (13 Tcf) in areas where the Whitehill Formation is at depths of more than 1500 m, contains less than 20% dolerite, and has a vitrinite reflectance of $<3.5\%$ (Figure 1).[2] Proving this resource is crucial, and a very important next step is testing it within an identified sweet spot, as is currently planned by the Council for Geoscience through drilling of an additional scientific stratigraphic borehole near Beaufort West.[38]

Acknowledgements

We acknowledge support from CIMERA which administered the KARIN funding, made possible through a donation by Shell. We thank the communities of the Tankwa Karoo and Willowvale for their support and ongoing interest. We thank Geoserve Exploration Drilling and Alec Birch for management of the drilling programme. We acknowledge the support staff of CIMERA (i.e. Ms Viwe Koti and Mr George Henry) for their contribution towards the execution of the project.

Authors' contributions

Project management and leadership was shared by M.O.d.K. and A.E.G., with significant contributions by N.J.B. A.E.G. was responsible for initial funding acquisition. M.O.d.K. wrote the initial draft, with all authors contributing during advanced stages of writing. E.O.A. is a master's student under supervision of M.O.d.K. and co-supervision of N.J.B. and F-G.O. Detailed lithostratigraphic logs of the CIMERA-KARIN cores were made by D.C. Borehole BH 47 was sampled by M.O.d.K., while sampling of the KZF-01 and KWV-01 was conducted by the KARIN researchers, including N.J.B., A.E.G., M.O.d.K., D.C. and C.G. Kübler index data were collected by E.O.A. and F-G.O., and vitrinite reflectance measurements were made by E.O.A. C.G. provided Rock-Eval data from elsewhere in the basin for comparison purposes.

References

1. Decker J, Marot J. Annexures: Investigation of hydraulic fracturing in the Karoo of South Africa. Pretoria: Department of Mineral Resources; 2012. Available from: http://www.dmr.gov.za/publications/viewdownload/182/854.html.

2. Cole D. Geology of Karoo shale gas and how this can influence economic gas recovery. Gas – The game changer for South Africa?! Paper presented at: Fossil Fuel Foundation Conference; 2014 May 12; Johannesburg, South Africa.

3. Kuuskraa V, Stevens S, Van Leeuwen T, Moodhe K. World shale gas and shale oil resource assessment. Technically recoverable shale oil and shale gas resources: An assessment of 137 shale formations in 41 countries outside the United States. Washington DC: US Department of Energy; 2013. Available from: http://www.eia.gov/analysis/studies/worldshalegas/pdf/fullreport.pdf.
4. Mowzer Z, Adams S. Shale gas prospectivity analysis of the southern Main Karoo basin. Petroleum Agency South Africa contribution to the strategic environmental assessment. Agency report FG 2015. Cape Town: Petroleum Agency South Africa; 2015.
5. Decker J. Geological evaluation of the Karoo basin's shale gas resource. Paper presented at: 4th Shale Gas Southern Africa Summit; 2014 March 24–26; Cape Town, South Africa.
6. Geel C, De Wit MJ, Booth PWK, Schulz H-M, Horsfield B. Palaeo-environment, diagenesis and characteristics of Permian black shales in the lower Karoo Supergroup flanking the Cape Fold Belt near Jansenville, Eastern Cape, South Africa: Implications for the shale gas potential of the Karoo basin. S Afr J Geol. 2015;118(3):249–274. http://dx.doi.org/10.2113/gssajg.118.3.249
7. Cole DI, Basson WA. Whitehill Formation. In: Johnson MR, editor. South African Committee for Stratigraphy Lithostratigraphic Series 3. Preoria: The Council for Geoscience; 1991. p. 51–53.
8. Kuuskraa V, Stevens S, Van Leeuwen T, Moodhe K. World shale gas resources: An initial assessment of 14 regions outside the United States. Washington DC: United States Energy Information Administration; 2011. Available from: http://www.eia.doe.gov/analysis/studies/worldshalegas.
9. Catuneanu O, Wopfner H, Eriksson PG, Cairncross B, Rubidge BS, Smith RMH, et al. The Karoo basins of south-central Africa. J Afr Earth Sci. 2005;43(1–3):211–253. http://dx.doi.org/10.1016/j.jafrearsci.2005.07.007
10. Johnson MR, Van Vuuren CJ, Hegenberger WF, Key R, Shoko U. Stratigraphy of the Karoo Supergroup in southern Africa: An overview. J Afr Earth Sci. 1996;23(1):3–15. http://dx.doi.org/10.1016/S0899-5362(96)00048-6
11. Jourdan F, Féraud G, Bertrand H, Kampunzu AB, Tshoso G, Watkeys MK, et al. Karoo large igneous province: Brevity, origin, and relation to mass extinction questioned by new 40Ar/39Ar age data. Geology. 2005;33(9):745. http://dx.doi.org/10.1130/g21632.1
12. Svensen H, Corfu F, Polteau S, Hammer Ø, Planke S. Rapid magma emplacement in the Karoo Large Igneous Province. Earth Planet Sci Lett. 2012;325–326:1–9. http://dx.doi.org/10.1016/j.epsl.2012.01.015
13. Hansma J, Tohver E, Schrank C, Jourdan F, Adams D. The timing of the Cape Orogeny: New 40Ar/39Ar age constraints on deformation and cooling of the Cape Fold Belt, South Africa. Gondwana Res. 2016;32:122–137. http://dx.doi.org/10.1016/j.gr.2015.02.005
14. Rowsell DM, De Swardt AMJ. Diagenesis in Cape and Karoo sediments, South Africa, and its bearing on their hydrocarbon potential. Trans Geol Soc S Afr. 1976;79:81–145.
15. Geel C, Schulz H-M, Booth P, De Wit M, Horsfield B. Shale gas characteristics of Permian black shales in South Africa: Results from recent drilling in the Ecca Group (Eastern Cape). Energy Procedia. 2013;40:256–265. http://dx.doi.org/10.1016/j.egypro.2013.08.030
16. Maré LP, De Kock MO, Cairncross B, Mouri H. Application of magnetic geothermometers in sedimentary basins: An example from the Western Karoo basin, South Africa. S Afr J Geol. 2014;117(1):1–14. http://dx.doi.org/10.2113/gssajg.117.1.1
17. Black DE, Booth PWK, De Wit MJ. Petrographic, geochemical and petro-physical analysis of the Collingham Formation near Jansenville, Eastern Cape, South Africa – Potential cap rocks to shale gas in the Karoo. S Afr J Geol. 2016;119(1):171–186. http://dx.doi.org/10.2113/gssajg.119.1.171
18. Smithard T, Bordy EM, Reid DL. The effect of dolerite intrusions on the hydrocarbon potential of the lower Permian Whitehill Formation (Karoo Supergroup) in South Africa and southern Namibia: A preliminary study. S Afr J Geol. 2015;118(4):489–510. http://dx.doi.org/10.2113/gssajg.118.4.489
19. Macey PH, Siegfried HP, Minaar H, Almond J, Botha PMW. The geology of Loeriesfontein area. Explanation: Sheet 3018. Pretoria: Council for Geoscience; 2011.
20. Van Vuuren CJ, Broad DS, Jungslager EHA, Roux J, McLachlan IR. Oil and gas. In: Wilson MGC, Anhaeusser C, editors. The mineral resources of South Africa. Pretoria: Council for Geoscience; 1998. p. 483–494.
21. Kingsley CS. Stratigraphy and sedimentology of the Ecca Group in the Eastern Cape Province, South Africa [PhD thesis]. Port Elizabeth: University of Port Elizabeth; 1977.
22. Götz AE, Ruckwied K. Palynology of the Permian Prince Albert and Whitehill Formations (Karoo basin, South Africa): New insights on basin dynamics and implications for shale gas exploration. Paper presented at: Joint meeting of TSOP-AASP-ICCP; 2016 Sep 18–23; Houston, Texas, USA.
23. Strauss H, Des Marais DJ, Hayes JM, Lambert IB, Summons RE. Procedures of whole rock and kerogen analysis. In: Schopf JW, Klein C, editors. The Proterozoic biosphere: A multidisciplinary study. Cambridge: Cambridge University Press; 1992. p. 699–707. https://doi.org/10.1017/CBO9780511601064.018
24. ASTM International. Standard test method for microscopical determination of reflectance of vitrinite dispersed in sedimentary rocks (D7708-14). West Conshohocken, PA: ASTM International; 2016 [cited 2017 Feb 15]. Available from: http://www.astm.org
25. Kübler B. Les argiles indicateurs de métamorphisme [The clay mineral indicators of metamorphism]. Revue Instituté de la Français de Pétrole. 1964;19:1093–1112. French.
26. Batchelor GK. An introduction to fluid dynamics. Cambridge: Cambridge University Press; 1967.
27. Lafargue E, Marquis F, Pillot D. Rock-Eval 6 applications in hydrocarbon exploration, production, and soil contamination studies. Revue de l'Institut Français du Pétrole. 1998;53(4):421–437. https://doi.org/10.2516/ogst:1998036
28. Espitalie J, Marquis F, Barsony I. Geochemical logging, Project B70 81018 Geology No. 25457. Paris: Institut Francais du Petrole; 1982.
29. Aarnes I, Svensen H, Polteau S, Planke S. Contact metamorphic devolatilization of shales in the Karoo basin, South Africa, and the effects of multiple sill intrusions. Chem Geol. 2011;281:181–194. http://dx.doi.org/10.1016/j.chemgeo.2010.12.007
30. Eberl DD, Velde B. Beyond the Kübler index. Clay Miner. 1989;24:571–577.
31. Dembicki H. Three common source rock evaluation errors made by geologists during prospect or play appraisals. Am Assoc Pet Geol Bull. 2009;93:341–356. http://dx.doi.org/10.1306/10230808076
32. Summons RE, Hope JM, Swart R, Walter MR. Origin of Nama basin bitumen seeps: Petroleum derived from a Permian lacustrine source rock traversing southwestern Gondwana. Org Geochem. 2008;39:589–607. http://dx.doi.org/10.1016/j.orggeochem.2007.12.002
33. Cole DI, McLachlan IR. Oil potential of the Permian Whitehill Shale Formation in the Main Karoo basin, South Africa. In: Ulbrich H, Rocha Campos AC, editors. Gondwana Seven Proceedings; 1988 July 18–22; Sao Paulo, Brazil. Sao Paulo: Instituto de Geociências, Universidade de Sao Paulo; 1991. p. 379–390.
34. Svensen H, Jamtvit B, Planke S, Chevallier L. Structure and evolution of hydrothermal vent complexes in the Karoo basin, South Africa. J Geol Soc London. 2006;163:671–682. https://doi.org/10.1144/1144-764905-037
35. Cole DI, Roberts DL. Other carbonaceous fuels. In: Wilson MGC, Anhaeusser C, editors. The mineral resources of South Africa. Pretoria: Council for Geoscience; 1998. p. 495–504.
36. Laughrey CD, Ruble TE, Lemmens H, Kostelnik J, Butcher AR, Walker G, et al. Black shale diagenesis: Insights from integrated high-definition analyses of post-mature Marcellus Formation rocks, Northern Pennsylvania. Paper presented at: American Association of Petroleum Geologists Annual Convention and Exhibition; 2011 April 10–13; Houston, Texas, USA.
37. Van der Spuy D. An overview of South Africa's oil and gas landscape. Proceedings of the 2nd SAMREF Meeting; 2016 October 27; Cape Town, South Africa. Cape Town: South African Marine Research and Exploration Forum; 2016.
38. Malumbazo N, Chevalier L. CGS stratigraphic borehole. Geobulletin. 2016;59(1):30–34.