South African Journal of Science

volume 114
number 5/6

Decolonising engineering in South Africa

The Grootfontein aquifer: Governance of a hydro-social system

Household food waste disposal in South Africa

South African behavioural ecology research in a global perspective

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Volume 114
Number 5/6
May/June 2018

South African Journal of Science

eISSN: 1996-7489

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
www.sajs.co.za

Cover caption
A meerkat (*Suricata suricatta*).
Meerkats feature prominently in
behavioural ecology studies, both locally and
internationally. In an article on page 21, Le Roux explores
the trends in behavioural ecology studies and places South African
research in a global perspective.

The Fourth Industrial Revolution and education

The inauguration of one of the world's leading specialists in artificial intelligence (AI) as the Vice Chancellor of a South African university has brought the Fourth Industrial Revolution to the fore in the local media – and raised interest in what the Fourth Industrial Revolution might mean for education in general – and for post-school education in particular. A fusion of technologies that is blurring the lines between the physical, digital and biological domains, AI's precise beginning is not clear, but it was certainly in evidence 3 years ago, if not earlier, building on the digital revolution.

The implications of the AI revolution for business, industry and daily life remain to some extent in the realm of speculation, but have nevertheless been discussed widely. Just what it might mean for education has had less attention, although the implications are extensive – both in terms of what universities can (or should) contribute to the advance of AI and its applications and how curricula and learning will need to change.

The most obvious matters are those that relate to the ways in which the nature of work and the job market are changing – and will continue to change at an increasing pace. It no longer makes sense to ask children what they would like to 'do' when they grow up. By the time they enter the world of work, a large portion of current job types will have disappeared, and as many (if not more) jobs, presently not defined, will have become both every day and essential. Apart from the nature of work, there is considerable disagreement as to whether or not the Fourth Industrial Revolution will create more employment or result in the loss of work opportunities.

Optimistic predictions, based on trends already measured, suggest that the next 3 years will see half a million more jobs created than those lost. As the global economy moves towards the widespread adoption of AI solutions, competition will grow for employees who have the scarce skills required to implement, manage and work alongside the new technology. Developing these skills is therefore vital for any young person wishing to remain relevant in an increasingly automated workplace.[1]

As this skilled workforce supports the AI industry, the demand for even more highly trained professionals will grow accordingly. Workplaces will require adaptable people whose jobs are reimagined, enriched or facilitated by the technology they work alongside.[1]

The pessimistic view is that AI will replace more jobs than it will create – with more than half of today's jobs becoming automated within the next 35 years. This view might, however, prove to be overly pessimistic. For example, the *Economist* reported recently that while computers can play chess better than almost all human beings, and help to diagnose diseases, excelling at abstract, cognitive tasks and at completing repetitive manual tasks, they are not yet as competent as people when it comes to undertaking non-repetitive physical tasks such as navigating randomly cluttered rooms or assembling pre-made furniture.[2] This will undoubtedly change in time, but it will be the pace of that change that determines the rate at which such tasks no longer need human engagement.

The implications that arise for institutions of education are at least twofold. Firstly, researchers in relevant disciplines face the challenge of making AI increasingly more sophisticated and useful, not just in manufacturing or planning but also in the direct service of society. The work of scientists, policymakers, social workers, educationists and many others whose duty of care it is to aim for the achievement of the 17 Sustainable Development Goals can all benefit from sophisticated AI applications. Whether the goal is quality education, decent work, climate action, affordable and clean energy or sustainable cities, there are already AI options of value and importance, yet more can and should be developed. But there are also other ways in which research (perhaps of a different nature) is important. In the realms of voice and facial recognition, for example, current systems are based on American and European norms, so that (for example) African or Chinese accents or facial features remain marginal. AI needs to be revolutionised and deracialised – and this requires research and interventions from scholars and scientists, a new activism, that goes beyond the creation of new algorithms.[3]

The second implication has to do with curricula, teaching and learning – rather than about robotic tutors. To succeed as a member of society, and as an employee, in the era of the Fourth Industrial Revolution, numeracy, literacy and an understanding of how the world operates are all essential. Students studying the basic and applied sciences need also to understand the political and social natures of the world in which they live. For the same reasons, students who study the humanities and social sciences need to understand at least the foundations on which AI is based and operates. This is a different kind of decolonisation of curricula – even requiring, perhaps, some of the elements of the kind of education provided (at least at first-year level) by liberal arts colleges.

The second implication has further requirements: people must have the skills required to implement, manage and work with the new technology, and with one another. And, not least, to be problem solvers, to be adaptable, and to be able to express themselves in both the written and spoken word – and to make the kinds of ethical and moral decisions that are not ever likely to become successful elements of AI. This challenge is one to which educators will have to rise.

References

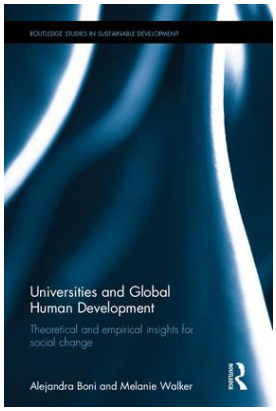
1. Marlin D. Millennials, this is how artificial intelligence will impact your job for better and worse. Forbes. 2018 January 16 [cited 2018 May 03]. Available from: https://www.forbes.com/sites/danielmarlin/2018/01/16/millennials-this-is-how-artificial-intelligence-will-impact-your-job-for-better-and-worse/#164ba5e44533
2. The Kamprad test: IKEA furniture and the limits of AI. The Economist. 2018 April 21; Leaders. Available from: https://www.economist.com/news/leaders/21740735-humans-have-had-good-run-most-recent-breakthrough-robotics-it-clear
3. Marwala T. Tackling bias in technology requires a new form of activism [webpage on the Internet]. c2018 [cited 2018 May 03]. Available from: https://www.uj.ac.za/newandevents/Pages/Opinion-Tackling-bias-in-technology-requires-a-new-form-of-activism.aspx

HOW TO CITE: Butler-Adam J. The Fourth Industrial Revolution and education. S Afr J Sci. 2018;114(5/6), Art. #a0271, 1 page. http://dx.doi.org/10.17159/sajs.2018/a0271

Towards human development friendly universities

BOOK TITLE:
Universities and global human development. Theoretical and empirical insights for social change

BOOK COVER:



AUTHORS:
Alejandra Boni and Melanie Walker

ISBN:
9780815355878 (softcover)

PUBLISHER:
Routledge, Abingdon; GBP36.99

PUBLISHED:
2016

REVIEWER:
Merridy Wilson-Strydom

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

EMAIL:
wilsonstrydommg@ufs.ac.za

HOW TO CITE:
Wilson-Strydom M. Towards human development friendly universities. S Afr J Sci. 2018;114(5/6), Art. #a0262, 2 pages. http://dx.doi.org/10.17159/sajs.2018/a0262

PUBLISHED:
30 May 2018

This book is timely and important for the higher education sector at local, national and global levels as well as for debates about what constitutes development and what role universities ought to have in development and social change. The authors propose a normative framework, based on human development ideas, that is of value for policymaking, scholarship and practitioners and that shows 'why higher education matters both in educational and social terms' (p. 6). Three objectives flow from this aim: (1) to deepen the link between universities and human development; (2) to operationalise human development in and through universities at a more local level; and (3) to encourage universities to question how and what knowledge is produced through research, and selected and mediated through teaching and learning within our universities. Key human development values such as participation, equity, sustainability, diversity and human security are emphasised throughout.

The book is presented in three parts. The first part sets out the relevant literature, conceptual framing, and operationalises the aims and objectives of the book. In the second part, the authors turn to empirical examples at both policy and practice levels and use these examples to question the nexus between higher education and human development. In the final part, these issues are woven together to make a 'robust but feasible case for universities, which could be for more rather than less human development and social justice in a world of growing inequalities' (p. 11).

Chapters 2 and 3 provide a helpful and critical overview of the role of universities when approached from an education and development perspective. For the reader new to the field of development studies or international development, the setting out of major development theories is useful. Chapter 4 provides a succinct and well-argued overview of the key concepts, ideas and implications of the human development and capability approaches. This chapter would be of particular value to those new to the area, but also serves as a helpful reminder of the key tenets of the approach for those who may be better versed in the ideas. The application to higher education points to the usefulness of the approach for thinking differently about universities, but also provides a helpful context within which to consider what human development ideas mean in practice.

The book then turns to policy and practice applications. These application chapters provide evidence of different ways of being and doing within universities and show why this is important. In the first application chapter, the authors analyse higher education policies from a human development point of view. Their approach to policy is a key strength of the chapter, with policy being understood in two senses: as texts (policy documents) produced by different actors and as processes of negotiation, contestation and struggle that may take place outside of the official policymaking machinery. Two quite different policy examples are interrogated – the Responsible Research and Innovation policy in the EU and Spanish context, and the 2013 White Paper for Post-School Education and Training in South Africa, with a particular focus on the section on quality.

Next, the authors turn to the knowledge and research contributions to human development that are made possible when universities pay attention to how research is conducted and the challenging epistemic questions of whose knowledge, for whose benefit and for whose capability expansion. The authors make the case not only for 'good research' – which is of course important – but instead for what they call 'good *for* research' – research that makes a difference in the world around us. Although less explicitly stated, the book also makes a case for 'good *for* higher education', as opposed to simply good higher education. These distinctions have critical implications for how we think about the purposes and quality of higher education. An important component of the argument in this chapter is the need to keep alive a commitment to university-produced knowledge as a public good, especially in a global context in which knowledge is typically approached as intellectual property not unlike other goods and services that are produced, accumulated and traded, often for financial gain.

In the following chapter, this argument becomes more concrete, with a specific focus on the potential of participatory action research (PAR) in the context of social sciences research. After setting out the key elements of PAR and reflecting on how this can support human development oriented research, Boni and Walker propose a new theoretical framework to analyse and inform PAR in higher education from a human development and capability approach point of view. In particular, they propose a 'participatory research capability cube' that has a tri-dimensional perspective, including: (1) a participatory axis that considers the expansion of the capabilities and agency of co-researchers, (2) a knowledge axis that takes account of the characteristics of the knowledge produced, and (3) a public deliberation axis that draws attention to the democratic processes that PAR can enable during and beyond a given research process. This innovative theoretical and analytical frame is then applied to understand a participatory research and teaching process that took place in the poor neighbourhood of San Lorenza in Castelon, Spain, as part of a 14-day intensive summer school for 30 master's degree students from across Europe.

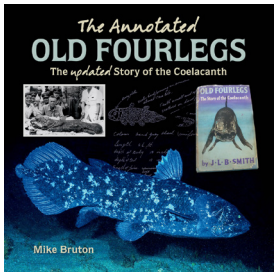
The final application chapter turns to student learning opportunities and outcomes and explores how this core function of the university can be understood from a human development perspective. The argument covers curriculum and pedagogy and highlights the importance of drilling down to the micro-space of learning conditions which are shaped by wider circumstances and inequalities in society. Three cases – two initiatives in Spain centred around curriculum and citizenship, and one in South Africa focused on employability – are used to illustrate the argument. The chapter shows that 'while neoliberal times may undoubtedly make life hard in universities, there are spaces to practice education differently…and to be agents as students and lecturers, rather than onlookers to policy' (p. 172).

In the last part of the book, Boni and Walker sum up the arguments and evidence presented across the chapters to make a case for what they call 'the human development friendly university'. This is a university that would be shaped by human development values in all dimensions of their core activities, including research, curriculum, pedagogy and policy. Although the community engagement function of the university is not explicitly considered in the book (a gap worth addressing in future work), the approach to research and to teaching and learning proposed implies meaningful engagement with actors outside of the university. To become a different type of university, a new language for thinking about universities and social change is needed. Boni and Walker propose that the human development approach could provide this different language, together with an alternative informational basis for assessing justice and social change contributions both within and beyond the university. The language of human development and capabilities enhancement provides a lexicon that could be used to challenge the current mainstream vision of a university that is based primarily on neo-liberal values. As has been shown in this book, there are exciting examples of research and teaching and learning practices that do take us in the direction of a human development friendly university – practices that meaningfully contribute to building more just and responsive universities, both within the university itself and with respect to the university's role in society.

This book is an important and challenging one that is well worth engaging with for those who wish to explore different purposes for higher education in society, and particularly for those who want to think and act differently across their various roles within the university.

A new look at Old Fourlegs

BOOK TITLE:
The annotated Old Fourlegs: The updated story of the coelacanth

BOOK COVER:


AUTHOR:
Mike Bruton

ISBN:
9781775844990 (softcover)

PUBLISHER:
Struik Nature, Cape Town; ZAR350

PUBLISHED:
2017

REVIEWER:
Brian W. van Wilgen

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

EMAIL:
bvanwilgen@sun.ac.za

HOW TO CITE:
Van Wilgen BW. A new look at Old Fourlegs. S Afr J Sci. 2018;114(5/6), Art. #a0269, 1 page. http://dx.doi.org/10.17159/sajs.2018/a0269

PUBLISHED:
30 May 2018

The discovery in 1938 of a coelacanth off the coast of East London was an event of immense importance to biological science. Up until then, coelacanths were only known from fossil records, and were believed to have been extinct for 66 million years. The discovery was akin to, for example, finding a pterodactyl flying around the Karoo. *Old Fourlegs* is a book, published in 1956, that recounts the story of the discovery of the first coelacanth, and the subsequent search, over 14 years, that led to the location and acquisition of a second specimen from the Comoro Islands. And what a dramatic tale it is. The fact that the first fish was captured at all is remarkable. A series of equally unlikely events followed. It was brought to the attention of the young local museum curator (Marjorie Courtenay-Latimer), who recognised it as highly unusual. She managed to convey this to a local chemistry lecturer and part-time ichthyologist (J.L.B. Smith) who was away on an end-of-year fishing holiday, and between them they managed to secure the specimen, so that it could later be adequately described. That these connections were made over the Christmas to New Year period, with everyone on holiday, and with the rudimentary communications technology of the 1930s, is almost miraculous. It changed the course of ichthyology in South Africa, and allowed J.L.B. Smith to focus on laying a solid foundation in that branch of science. The story of the location and retrieval of the second coelacanth 14 years later is even more dramatic. It too was discovered in a remote place (Anjouan Island in the Comoros) at the end of the year, when everyone was on holiday. Smith managed to locate and contact the Prime Minister of South Africa (D.F. Malan), and persuade him to place an Air Force Dakota and crew at his disposal to retrieve the specimen. They proceeded, essentially unannounced, across Portuguese East Africa to the Comoros, where they snatched the specimen from under the noses of the French government, and brought it back to South Africa. It is one of the most exciting scientific adventures I have ever read.

Mike Bruton's publication is a facsimile reproduction of the 1956 first edition of *Old Fourlegs*, with (as the title suggests) many anecdotes and updates. In itself, *Old Fourlegs* is a remarkable book. Following its publication in 1956, *Old Fourlegs* rapidly became an international best-seller and one of the most popular scientific books worldwide at the time. It has appeared in six further English editions, and has been translated into French, German, Russian, Estonian, Afrikaans, Dutch, Slovak, Latvian and Japanese. As a student of J.L.B. Smith, and having gone on to become Director of the (then) J.L.B. Smith Institute of Ichthyology, Mike Bruton is exceptionally well placed to compile the anecdotes and updates.

The annotations themselves, and there are many, provide fascinating insights as well as trivia (e.g. the nickname for the Dakota that was dispatched to collect the coelacanth from the Comoros was the 'Flying Fishcart', and it is now in the South African Air Force Museum at Ysterplaat). The characters involved are also fascinating. Anyone who reads the original book will realise that J.L.B. Smith had little time for J.C. Smuts, a prime minister who had refused him help at one stage, but a lot of respect for D.F. Malan, who did help him. The difference of approach between these two politicians also seems out of character, as Smuts was a scientist and a liberal, while Malan was a conservative theologian and creationist. That help came from the latter and not the former was therefore surprising. Mike Bruton speculates that because Smuts' rapport with nature was rooted in botany and mountains, he probably 'lacked empathy for aquatic life, as if rivers and oceans were on another planet'. Unlike Malan, Smuts was also inclined to accord more respect to foreign scientists than to scientists (and other experts) from his own country. Smith placed his second coelacanth in the genus *Malania* out of gratitude to Malan, although it was later confirmed that it was in fact the same species as the first one. Another detail that I had not noticed before concerns Smith's monumental work *The Sea Fishes of Southern Africa*, which first appeared in 1949, and has been updated and re-published many times. The more recent titles were *Smith's Sea Fishes* (1977) and *Smiths' Sea Fishes* (1986), and our attention is brought to the subtle shifting of the apostrophe in order to correctly recognise that the second work embodies the substantial contributions of J.L.B. Smith's wife Margaret both as an illustrator and a fish taxonomist. Today we know that there are two species of living coelacanths – the original *Latimeria chalumnae* (widespread in the western Indian Ocean) and *L. menadoensis* (discovered in a fish market in Indonesia in 1997). Interestingly, no coelacanth has ever been caught on purpose by Western scientists, despite multimillion dollar efforts to do so.

If you have not yet read the original *Old Fourlegs*, then obtaining a copy of Bruton's book is worth it for that reason alone. But this book is a whole lot more. There are notes and photographs next to each reproduced page, each providing additional insights, updates, or comparisons of modern understanding to predictions put forward in 1956. Finally, the book also contains a biography of J.L.B. Smith, a publication history of *Old Fourlegs*, as well as of other books on coelacanths, and other books written by J.L.B. Smith. In addition, there is a chapter that provides an overview of coelacanth discoveries, coelacanth biology, and coelacanth conservation, and a further chapter on the significance of the coelacanth in science, culture and art. I would recommend this book unreservedly to anyone with an interest in fish, history or adventure.

Decolonising engineering in South Africa – Experience to date and some emerging challenges

AUTHOR:
Mike Muller[1]

AFFILIATION:
[1]Visiting Adjunct Professor, Wits School of Governance, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Mike Muller

EMAIL:
mike.muller@wits.ac.za

KEYWORDS:
decolonisation; developmental state; water management

HOW TO CITE:
Muller M. Decolonising engineering in South Africa – Experience to date and some emerging challenges. S Afr J Sci. 2018;114(5/6), Art. #a0270, 6 pages. http://dx.doi.org/10.17159/sajs.2018/a0270

PUBLISHED:
30 May 2018

Decolonisation is poorly defined and contentious, particularly when applied in the domain of 'pure' sciences.[1] However, engineering involves the application of science in society and the political and cultural context has obvious relevance. Here, the implications of the decolonisation discourse for engineering are considered, using the lens of water – an archetypal focus of public policy and management.

This consideration arose from an invitation to present the Annual Lecture of the South African Academy of Engineering (SAAE) which seeks 'To be a reliable and sought-after source of expert advice on matters pertaining to global competitiveness and quality of life for the nation'[2]. The lecture was given on three university campuses, and was tailored to local issues to promote discussion. This discussion provided useful insights into the meaning and relevance of decolonisation in the engineering domain.

Two propositions informed the approach. First, was that engineering and the applied sciences are about translating knowledge into action to achieve practical goals. There is wide consensus on the societal goals of 'water security'[3] which avoids a debate about by whom and how goals are determined. The second proposition distinguishes between retrospective approaches which deconstruct colonial norms and values to understand their impact on the present and prospective approaches which consider what to do next. Should the priority be to break down the old or simply to recognise the flawed foundation and build a durable new future, perhaps through a developmental state tasked to achieve a new national vision?

The cases

In Cape Town, the focus was on the self-styled 'water crisis' whose origin and characterisation, it is suggested, is linked to the continued influence of European approaches inappropriate to African challenges.

The question in Port Elizabeth was whether a decolonised future will be characterised by new elites simply capturing the privileges of their predecessors? – the neo-colonialism described by writers such as Frantz Fanon[5].

In Johannesburg, government has put the city at risk of a water crisis by delaying investments in water security to support 'transformation'. Was it appropriate to undermine national development goals and allow a 'recolonisation' of knowledge?

The colonial history

To start, it was acknowledged that engineering was part of the colonial project. The British Institution of Civil Engineers was established as:

> *A Society for the general advancement of Mechanical Science, and more particularly for promoting the acquisition of that species of knowledge which constitutes the profession of a Civil Engineer; being the art of directing the great sources of power in Nature for the use and convenience of man …. as the means of production and of traffic in states, both for external and internal trade, as applied in the construction of roads, bridges, aqueducts, canals, river navigation, and docks, for internal intercourse and exchange; and in the construction of ports, harbours, moles, breakwaters, and lighthouses, and in the art of navigation by artificial power, for the purposes of commerce; ….*[and] *in the drainage of cities and towns.*[6]

So civil engineering was about means of production, trade and commerce. It was the engineers who built the harbours and the roads along which the occupying forces advanced and who laid the tracks of the railways that enabled the metropoles to enlarge their wealth by extracting the raw materials of colonised countries.

The first civil engineering contract in South Africa was for a canal in Cape Town to supply fresh water to passing ships of the Dutch East India Company. The profession's history is of the roads, mountain passes and railways, the dams, power stations and transmission lines that enabled settlement and made mining possible. Even water and sanitation was provided mainly for the colonisers, leaving a legacy of separate standards for white and black, rural and urban. Meanwhile, because black people were excluded, some stigma remains: can the profession be trusted to serve the interests of the majority?

Colonised engineering: The experience of Arthur Lewis

Colonialism was not just about economic infrastructure and colonial policy but had direct impacts on people. Arthur Lewis, a great academic from the African diaspora, won the Nobel Prize for economics in 1979 for his work on growth and employment in Africa. As he explained in his Nobel acceptance biography[7], he originally wanted to be an engineer:

> *In 1932 I sat the examination and won the scholarship. At this point I did not know what to do with my life. The British government imposed a colour bar in its colonies, so young blacks went in only for law or medicine where they could make a living without government support. I did not want to be a lawyer or a doctor. I wanted to be an engineer, but this seemed pointless since neither the government nor the white firms would employ a black engineer.*

Lewis's admirably concise description of his life choices illustrates simply how the wider discrimination in the European colonies at that time was reflected in engineering.

Cape Town's colonial approach

Cape Town's descent into water crisis, described elsewhere[8], shows how colonial influence persists even after formal political change. At the least, the City's approach was excessively influenced by a Eurocentric paradigm, reflected in Europe's Water Framework Directive[9], that seeks to avoid new infrastructure investments.

The City's leadership trusted this European paradigm rather than the tested systems approaches that have kept urban South Africa water secure.[10] They delayed new water supply investment because they believed that their conservation programmes could sufficiently curtail demand. This had happened before around 2000, when Cape Town's environmental community opposed construction of the Berg River Dam. It was not needed they said – conservation and alien plant clearance would suffice.[11]

On that occasion, a minor drought emphasised the need for action; without that dam, the City would have come much closer to its 'Day Zero' at the start of 2018. Yet, in 2013, city decision-makers once again stated that new infrastructure would not be needed before 2022/2024, despite recommendations of national government and the Planning Commission. They were convinced that it was their (excellent) conservation programmes rather than 3 years of good rains that had reduced consumption.

Two dry years and one year of drought later, supply restrictions were imposed. Post-hoc, this was blamed on severe drought (citing rainfall records[12] not representative of the catchments concerned nor acknowledging more nuanced South African Weather Service accounts[13,14]). Areal rainfall summary maps (Figure 1), while coarse grained, show rainfall at at least 75% of average in 2014/2015; between 50% and 150% in 2015/2016; and only reducing to 50–75% in 2016/2017.[15] Streamflow records from the catchment areas of the major dams also present a less dramatic picture.[16]

European environmentalists in countries with temperate climates, a substantial endowment of old infrastructure, stable populations and rich economies question the need for new infrastructure. But codified conservation is grossly inappropriate in African countries with much higher population, economic and urban growth all driving increased water use.[17,18]

Nelson Mandela Bay Metro and the post-colonial predatory state

In analyses of post-colonial development in Africa, concepts of the 'predatory state' and 'neopatrimonialism' figure large, suggesting that new power elites capture the resources of their societies for their own benefit. Replacing one set of exploitative elites by another does not constitute decolonisation but the danger is that it gives rise to a 'deterministically pessimistic view of development in Africa'[19].

Water shortages in Nelson Mandela Bay Metropolitan Municipality in 2017 and 2018 did not occur because of water resource constraints. Apartheid's hydraulic empire building phase in the 1960s saw the Orange River Scheme built, principally to benefit the farming constituency but also to provide water to Port Elizabeth,[20] which still has an unused allocation. Yields from Port Elizabeth's old dams were inadequate and

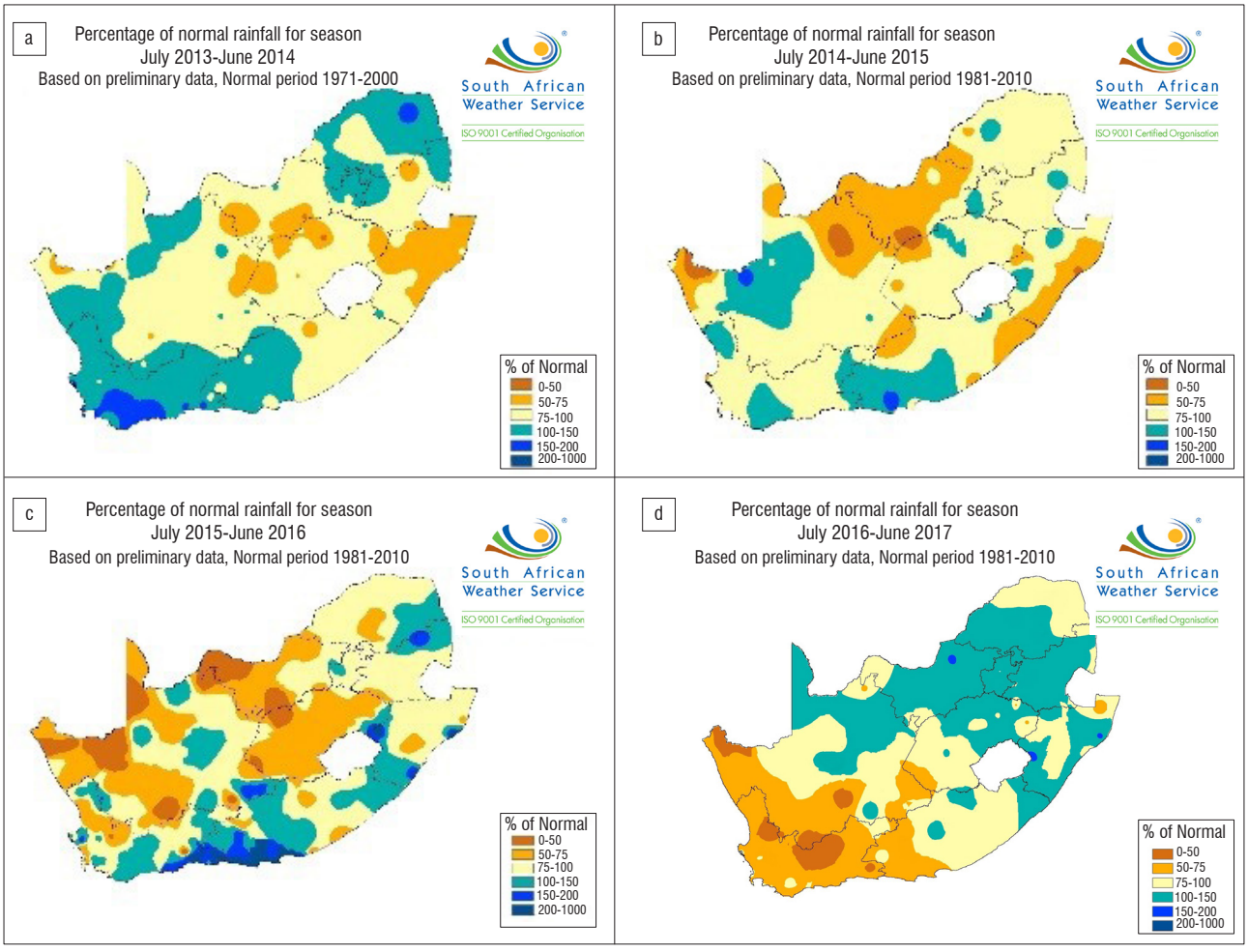


Figure 1: Areal rainfall summary maps: (a) 2013/2014, (b) 2014/2015, (c) 2015/2016 and (d) 2016/2017.

the Nooitgedacht works, long identified as the next supplementation for the city, were incomplete although the City's 2006 water plan stated that they would be needed by 2015, confirmed by the Algoa Supply System analysis in 2011.

The project was repeatedly delayed because a very expensive desalination option was preferred. Around 2004, there were appeals to national government to fund a project as a drought emergency. A new company would produce salt, chemicals and desalinated water; in 2010 progress appeared 'imminent' but in 2015, the company was deregistered. The Municipality is infamous for mismanagement and the fact that public investments only proceed if local political elites benefit.[21] The Nooitgedacht scheme was finally started, although it is not yet complete and further investments within the City will be needed. It has taken over 10 years to build a project that should have been completed in 3 years.

This raises sharp questions about the role of engineers in a 'post-colonial' South Africa. Should they simply serve new leaderships and build whatever they are told, regardless of better options? This lived reality of many government technicians surely does not represent decolonisation.

Planning the Vaal System – shallow transformation invites recolonisation

South Africa is correctly committed to transforming the demographics of its institutions – a process often treated as synonymous with decolonisation. This transformation faced challenges in technical institutions, given the limited pool of expertise initially available. The impact of this challenge on water security in the Gauteng economic heartland offers another perspective on decolonisation.

The transformation of the national water department included efforts to increase participation of black-owned companies in its business. But is transformation or decolonisation achieved merely by changing demographics? Or is the objective to ensure substantive participation by black professionals as both clients and service providers?

Water security for South Africa's cities is planned using complex systems modelling, developed in the 1970/1980s when the computing power needed first became available.[22] Models estimate the yield of interconnected systems as a whole, not just the sum of individual sources. Reliability levels are determined by stochastic methods, using rainfall and run-off variability data to generate flow and storage sequences. This approach has successfully informed operations, users and decision-makers.[10]

Recommendations derived from the models are not always followed, as Cape Town and Nelson Mandela Bay Metro Municipalities have demonstrated. But they have underpinned three decades of water security in the more complex Vaal River System, despite serious drought challenges.

This is specialised work, undertaken by a handful of consultants who have developed the necessary expertise. However, in 2014, the new Minister of Water explicitly sought to change this approach. She refused to appoint 'the same old companies', saying that 'she would decide which projects to build and who would build them'[23]. She introduced a 'panel system' (often excluding qualified companies) and awarded tenders only to preferred panel members.

This procurement system created lucrative (and sometimes corrupt) consulting opportunities on large projects but also applied to the systems modelling firms. When the Department's financial management collapsed, amidst allegations of systemic corruption and mismanagement,[24,25] funds for planning and associated modelling were slashed; when work restarted, experienced firms found that they were now expected to work as sub-contractors to 'panel' members. The new firms had limited capabilities but the old firms now had limited funds to train new staff. Meanwhile, with just a handful of skilled staff left, the Department no longer offers new graduates supervised technical experience – its historical training function.

This process has not only weakened oversight of water security in the big cities but has also blocked the reproduction of skills in a field in which South Africa had been a global leader. Local firms are being taken over by foreign companies that often use external resources for specialised technical work. Breaking down old institutions rather than building new ones is effectively ceding South Africa's capacities to foreign interests. Rather than decolonising, South Africa is inviting recolonisation: a process aggravated by weak strategic management and sometimes motivated by corruption.

Discussion

Short-term view and other problems of politicians

In the public presentations, the divide drawn between engineers and politicians raised some discomfort. Yet it was acknowledged that, in the cases presented, politicians' actions were not in the public interest. Damage was done when technical recommendations were not implemented or deliberately ignored in pursuit of private interests.

There is a wider debate over the roles that politicians and technocrats should play in public administration. Few politicians have the skills to make technical judgements about the work they oversee. Yet, since 1994, they have taken on increasingly operational roles rather than concentrating on oversight. This trend has led to problems when their priorities have deviated from formal goals. The situation has been aggravated by the appointment of malleable, rather than technically competent, officials, which weakens institutional capacity and the quality of advice.

Engineers and politicians work to different time scales. Engineers often plan decades ahead. Meanwhile, for politicians, short term means this week, and long term is until the next election. This short horizon inevitably shifts priorities. Politicians also need to mobilise political support and build constituencies. The obvious temptation is to use patronage, allocating resources in return for political and financial support. While public administrators have similar temptations, a well-functioning system ensures oversight; political heads appoint competent people to ensure that procedures are followed and goals are achieved. When politicians *take* management decisions, this oversight is lost, demonstrating that, as Fanon said, 'an engineer is a thousand times more indispensable to his country than an officer'[25].

While some dramatic cases of the resultant water sector corruption have been chronicled, more important issues are often ignored. Expensive projects – poorly conceived, implemented and operated – are failing to meet their objectives; the reliability of water supplies is decreasing.[26,27] The water users who suffer most are those living furthest from the cities. Even when it works, over-priced infrastructure is a liability not an asset, which increases costs of living and doing business and contributes to the financial crisis of many municipalities.

The role of the engineers is also challenged

But what are the expectations of technical professionals such as engineers? They must deal with the contradictions in their role evident in dictionary definitions of professionals as either[28]:

> *1: engaged in one of the learned professions, characterized by or conforming to the technical or ethical standards of a profession, exhibiting a courteous, conscientious, and generally business-like manner in the workplace; or*
>
> *2: participating for gain or livelihood in an activity or field of endeavour often engaged in by amateurs; engaged in by persons receiving financial return.*

Engineers might like the first definition, but trust in experts has declined, with some justification. Accountants present information that is false or misleading; doctors manipulate research results to promote dubious medicines. These are just hired hands, promoting commercial interests, not ethical 'professionals'. But healthy scepticism is being nurtured, a

syndrome of which Donald Trump is just a symptom. When, in Cape Town, academics and activists assert that water conservation will meet new water demands, they can easily reject contrary views from hydrologists and engineers as just another group promoting its own interests. With new factors like climate change, technical predictions of how much water is reliably available can be challenged, simply by saying that, surely, 'this time it's different!'.

The role of professionals in a developmental state

A 'developmental state' approach is often promoted as the way to address the country's structural challenges. In traditional models, technocrats enjoy a status of 'embedded autonomy'. While part of the administration, they are insulated from politicians who seek political or personal gain rather than technical goals.[29]

In what is characterised as South Africa's apartheid 'developmental state'[30], the core cadre of engineers and scientists were produced by the state. This followed recognition of excessive investment in the Orange River Scheme and the need for professional water management to avoid water becoming a brake on national development. A Commission of Enquiry analysed water management policies and priorities, and produced a comprehensive guide to the future.[31]

A key recommendation was to professionalise water management. Investment in water research and student bursaries created a community of professionals who could cooperate around common goals. They enjoyed a degree of trust that gave them considerable scope to implement and innovate – an environment in which it was possible to develop new approaches and undertake complex and challenging projects.[10,32] It is this 'production line' that has been lost as the state's capacity to absorb and train new graduates has weakened and most practical work is outsourced. This situation has also broken the mechanisms that integrated these professionals into society as 'trusted experts'. That breakdown has been aggravated by suggestion that professional organisations are closed clubs, seeking to maintain privileges and control entry. This global challenge is particularly acute in South Africa where a white old guard apparently controls the entry of the new cadre of young black engineers.

Society decides what tasks to delegate to their 'technicians' to protect the public. Whatever the approach, the concern for engineers is that their advice should be accepted and trusted. The challenge for technical professionals in 21st-century South Africa is to recognise that trust in 'experts' derives as much from their embeddedness in their society as from their technical ability.

This statement reinforces a general finding on water management: water problems are primarily people problems or, more precisely, socio-political rather than technical. Technical pathways to water security are often complex and long term. Professionals such as engineers in a developmental state must help politicians, administrations and the communities they serve to make and implement decisions when they are needed. To do that, they need to be trusted.

Conclusions: Decolonisation means asserting new roles

While the past obviously influences the future, it is useful to maintain separate perspectives, to avoid the trap of path dependence. Otherwise, if future strategies are primarily a response to the past, the past will continue to determine the future.

So, yes, engineering has undoubtedly been coloured by colonial objectives and attitudes that determined who came into the profession and what they could do; and equally important, who was excluded, with what consequences. Its purpose was, to a greater or lesser extent, to advance the colonial mission. Indeed, from the vantage point of economics, Lewis insisted that resource colonialism, the beginning of the new international economic order, began as '…. an off-shoot of the transport revolutions. The railway was a major element here'[33].

Because of this history, many engineering institutions are perceived as old fashioned and conservative. This perception is not helpful because engineers work in a world in which the role of technical expertise is increasingly challenged even as it becomes more vital. Engineers are vulnerable in this situation because their solutions to apparently simple challenges – like ensuring reliable water supplies – are often long term, complex and not understood or effectively communicated. This makes their recommendations easier to reject or just ignore.

Cape Town's plight reflects this situation. Its decision-makers were not telling lies when they claimed that they had permanently reduced water consumption. Their mistake was to claim an easy victory for a focus on demand management, ignoring more obvious drivers such as weather and population growth, and options such as increasing supply. But engineers failed too. They were not heard promoting an alternative, cautionary view because they were not adequately 'embedded' in that society, distrustful as it is of national politicians and solutions that might damage 'nature'.

The failure to date to articulate a clear vision of a decolonised future is unfortunate because it leaves many negative tendencies unchallenged. One example is the undermining of local capacity to model water resource systems; such capacity is critical to support both planning and operations. Undermining local capacity is opening the way for a recolonisation of a technical domain in which South Africa until recently had sovereign leadership.

While South African society will not be well served if it simply ignores engineers and their institutions, the responsibility also lies with the profession itself. Continued introspection is needed, to acknowledge and address its problematic history and legacy. But priority must be given to the new forces that are shaping the future.

The composition of the profession is changing radically, with a growing cadre of young black participants, including many women (Figure 2).[34] Over the next decade, the well-recognised gap of experienced 'middle professionals' will be filled. Many of today's young engineers who have worked with little support will turn out to have absorbed a great deal of wisdom as they navigated difficult circumstances.

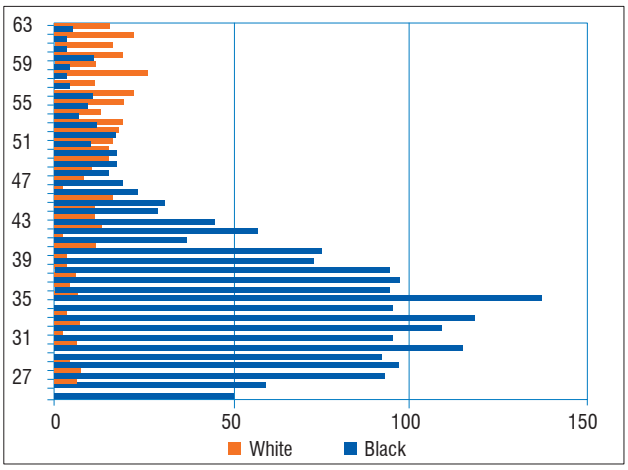


Figure 2: Engineering staff in (local) government, by race in 2015.

This change will not automatically transform other dimensions of the profession. Engineers will still be expected to exhibit those qualities of objectivity, ethical behaviour as well as technical competence inherent in the word 'professional'. They will need fortitude – and appropriate mechanisms – to manage the risks encountered in a competitive and acquisitive economy and polity. But the task will become easier as the complexion of the profession changes.

Buoyed by Fanon's estimate of their value, engineers should be more assertive in promoting the vital role that technical professionals must play if South Africa is to move towards the vision of a developmental state. Rather than being heard as voices from the past, they will be seen as pathfinders of the future, explaining what needs to be done. That is, in large measure, what decolonisation should be about.

The new generation of engineers will still need to fight for the space to do its work. In the water sector, they will have to be vocal in proclaiming the goal of water security, providing a sustainable and reliable water supply and sanitation services for people and economic activity. That will help them when they challenge politicians on the governance of technically focused public institutions. They must demand a clear distinction between what technical planners and managers do and what political heads *should* do. Engineers can engage, analyse, prioritise, and make and implement recommendations efficiently and effectively. Politicians should ensure that this is done competently, within clear policy frameworks, and following procedures that protect the public from attempts to make personal gains.

A further challenge will be to engineer effective institutions and systems. Cape Town's tribulations have helpfully shown that a failure to use technical expertise effectively is not the province of just one political party. All politicians should beware the arrogance of ignorance and learn to nurture and harvest sound advice, not ignore it. The difficulties emerging in Vaal System planning show that transforming institutions and decolonising knowledge is not just about demographics.

These challenges are most obvious for engineers in the public sector, because they are responsible for basic structures and services that are the foundations of society. But similar challenges face practitioners in other fields of engineering – mining, chemical, electrical and mechanical. Their work can transform conditions in the larger society for better or worse. They too must do more to make the voices of professional engineers heard, as critical contributors to public debate who inform the decisions of public and private policymakers.

This process will continue, and be contested, as long as society develops and evolves. Meanwhile, in the wider community, people will continue to measure society's progress by turning on the taps in their homes to see whether they can take at least this one of life's basic needs for granted. For its part, water will continue to teach that the reliability of its flows is determined by people, not hydrology. It will flow when and where needed as a result of the right decisions taken at the right time, informed by the best possible advice. This will often come in large measure from a truly decolonised cadre of engineers that is respected and nurtured by the communities in which it works.

References

1. Nordling L. How decolonization could reshape South African science. Nature. 2018;554:159–162. https://doi.org/10.1038/d41586-018-01696-w
2. South African Academy of Engineering (SAAE). Mission and vision [webpage on the Internet]. No date [cited 2018 May 07]. Available from: http://saae.co.za/about-us/
3. Grey D, Sadoff C. Sink or swim – Water security for growth and development. Water Pol. 2007;9(6):545–571. https://doi.org/10.2166/wp.2007.021
4. National Planning Commission (NPC). National development plan. Pretoria: Presidency; 2012.
5. Fanon F. The wretched of the earth [translated]. London: MacGibbon and Kee; 1965.
6. Institution of Civil Engineers. Royal charter, by-laws, regulations and rules. London: Institution of Civil Engineers; 2015. Available from: https://www.ice.org.uk/ICEDevelopmentWebPortal/media/Documents/About%20Us/Royal-Charter-By-laws-Regulations-and-Rules-2015.pdf
7. Lewis WA. Sir Arthur Lewis – Biographical [webpage on the Internet]. c2014 [cited 2018 May 07]. Available from: https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1979/lewis-bio.html
8. Muller M. Understanding the origins of Cape Town's water crisis. Civil Engineering. 2017;25(5):11–16.
9. European Commission. Water framework directive 2000 [webpage on the Internet]. c2000 [cited 2018 May 07]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32000L0060
10. Basson MS, Van Rooyen JA. Practical application of probabilistic approaches to the management of water resource systems. J Hydrol. 2001;241(1–2):53–61. https://doi.org/10.1016/S0022-1694(00)00367-X
11. Thompson L. Managing mobilisation? Participatory processes and dam building in South Africa, the Berg River Project. Working Paper Series 254. Brighton: IDS; 2005. Available from: http://opendocs.ids.ac.uk/opendocs/handle/123456789/4047
12. Wolski P. Facts are few, opinions plenty… on drought severity again (blog post on the Internet]. c2018 [cited 2018 May 07]. Available from: http://www.csag.uct.ac.za/2018/01/22/facts-are-few-opinions-plenty-on-drought-severity-again/
13. Kruger AC, Nxumalo MP. Historical rainfall trends in South Africa: 1921-2015. Water SA. 2017;43(2):285–297. https://doi.org/10.4314/wsa.v43i2.12
14. South African Weather Service (SAWS). Media releases of 26 January [webpage on the Internet]. c2018 [cited 2018 May 07]. Available from: http://www.weathersa.co.za/news-events/media-release
15. South African Weather Service (SAWS). Historical rain maps [webpage on the Internet]. No date [cited 2018 May 07]. Available from: http://www.weathersa.co.za/climate/historical-rain-maps
16. South African Department of Water and Sanitation (DWS). Hydrology: drainage regions G & H [webpage on the Internet]. No date [cited 2018 May 07]. Available from: https://www.dwa.gov.za/Hydrology/Verified/hymain.aspx
17. United Nations Department of Economic and Social Affairs (UNDESA). World urbanization prospects: The 2014 Revision: Highlights. Report ST/ESA/SER.A/352. New York: Population Division, UNDESA; 2014.
18. European Commission. Nature-based solutions [webpage on the Internet]. No date [cited 2018 May 07]. Available from: https://ec.europa.eu/research/environment/index.cfm?pg=nbs
19. Mkandawire T. Neopatrimonialism and the political economy of economic performance in Africa: Critical reflections. World Politics. 2015;67(3):563–612. https://doi.org/10.1017/S004388711500009X
20. South African Department of Water Affairs and Forestry (DWAF). History of the Orange River Project [webpage on the Internet]. No date [cited 2018 May 07]. Available from: http://www.dwaf.gov.za/orange/Mid_Orange/overview.htm
21. Olver C. How to steal a city: The Battle for Nelson Mandela Bay. Johannesburg: Jonathan Ball; 2017.
22. Basson MS, Allen RB, Pegram GGS, Van Rooyen JA. Probabilistic management of water resource and hydropower systems. Highlands Ranch, CO: Water Resources Publications; 1994.
23. Blom N. Statist ideology bars water crisis solution. Business Day. 2017 December 12; National/Science & Environment [cited 2018 May 07]. Available from: https://www.businesslive.co.za/bd/national/science-and-environment/2017-12-12-statist-ideology-bars-water-crisis-solution/
24. Auditor General South Africa (AGSA). Report of the Auditor-General to the Joint Committee of Inquiry into the functioning of the Department of Water and Sanitation: Challenges facing the water and sanitation portfolio. Pretoria: AGSA; 2018.
25. Davis R, Payne S. Gugile Nkwinti inherits a Water Department that may tip SA into chaos. Daily Maverick. 2018 February 27. Available from: https://www.dailymaverick.co.za/article/2018-02-27-cabinet-reshuffle-gugile-nkwinti-inherits-a-water-department-that-may-tip-sa-into-chaos/#.WvLw6oiFOUl
26. Statistics South Africa (StatsSA). Community survey 2016 data. Pretoria: StatsSA; 2016.
27. Statistics South Africa (StatsSA). GHS series report volume VIII. Water and sanitation: In-depth analysis of the General Household Survey 2002–2015. Pretoria: StatsSA; 2016.
28. Merriam-Webster Dictionary [online]. Professional [updated 2018 Apr 30; cited 2018 May 07]. Available from: https://www.merriam-webster.com/dictionary/professional
29. Evans P. Government action, social capital and development: Reviewing the evidence on synergy. World Develop. 1996;24(6):1119–1132. https://doi.org/10.1016/0305-750X(96)00021-6
30. Freund B. A ghost from the past: The South African developmental state of the 1940s. Transformation. 2013;81(1):86–114. https://doi.org/10.1353/trn.2013.0007

31. South Africa. Report of the Commission of Enquiry into Water Matters. R.P.34/1970. Pretoria: Government Printing Works; 1970.

32. Muller M. Greater security with less water: Sterkfontein Dam's contribution to systemic resilience. In: Increasing resilience to climate variability and change. Singapore: Springer; 2016. p. 251–278. https://doi.org/10.1007/978-981-10-1914-2_12

33. Lewis WA. The evolution of the international economic order. Discussion paper number 74. Princeton, NJ: Research Programme in Development Studies, Woodrow Wilson School, Princeton University; 1977.

34. Lawless A. Numbers and needs in local government – update 2015. Paper presented at: The Annual Conference of the Institute of Municipal Engineering of Southern Africa; 2016 October 26–28; East London; South Africa.

Response to Benoit and Thackeray (2017): 'A cladistic analysis of *Graecopithecus*'

AUTHORS:
Jochen Fuss[1]
Nikolai Spassov[2]
Madelaine Böhme[3]
David R. Begun[4]

AFFILIATIONS:
[1]Senckenberg Centre for Human Evolution and Palaeoenvironment, Tübingen, Germany
[2]National Museum of Natural History, Bulgarian Academy of Sciences, Sofia, Bulgaria
[3]Department of Geoscience, Eberhard Karls University of Tübingen, Tübingen, Germany
[4]Department of Anthropology, University of Toronto, Toronto, Ontario, Canada

CORRESPONDENCE TO:
Jochen Fuss

EMAIL:
jochen.fuss@uni-tuebingen.de

KEYWORDS:
Hominini; Africa; Eurasia

HOW TO CITE:
Fuss J, Spassov N, Böhme M, Begun DR. Response to Benoit and Thackeray (2017): 'A cladistic analysis of *Graecopithecus*'. S Afr J Sci. 2018;114(5/6), Art. #a0267, 2 pages. http://dx.doi.org/10.17159/sajs.2018/a0267

PUBLISHED:
30 May 2018

Benoit and Thackeray[1] provide a cladistic analysis that aims to refute the hypotheses that *Graecopithecus* is a member of the hominin clade and that hominins could have originated in the Eastern Mediterranean. In our response, we point out that the authors' thesis relies on a selective use of data and a series of misrepresentations of our results and conclusions that reflect what we see as an a priori hostility to the very idea of a non-African origin of the hominin clade.

It is useful to recall that ever since the widespread acceptance among scientists of the reality of human evolution there has been debate, often contentious, about the place of origin of the first humans (hominins). While Darwin and Huxley advocated Africa, Darwin recognised the possibility that it might be Europe – a fact often overlooked today. Haeckel and Dubois believed it was Asia while Dawson profited from the racist ideology of the day to promote a forgery (*Eoanthropus*, a.k.a. Piltdown) as proof that it was Europe. Osborn even proposed Nebraska! Following the description by Dart of *Australopithecus,* and especially the subsequent discoveries by Broom, the overwhelming consensus has been that Africa is the continent of hominin origins. This has become so widely accepted as to rise to the level of dogma, with the result that any claim to the contrary is automatically disputed if not ridiculed. The media coverage to which Benoit and Thackeray refer in their comment concerning the hypothesis of a European origin of hominins – when in fact we propose an Eastern Mediterranean (which also includes Africa) origin – is a perfect example.

Fuss et al.[2] propose that *Graecopithecus* may be a hominin, given the small size of the root of the lower canine but especially the root morphology of the lower premolar in the Greek specimen and the upper premolar in the specimen from Bulgaria (not lower premolar, contra Benoit and Thackeray). We note that if it is a hominin, *Graecopithecus* would be the oldest known. Given what we know about mammalian faunas in the Eastern Mediterranean and Africa between 10 and 7 Ma, dispersals clearly occurred between the two areas and, as is the case for elephantids, giraffids and bovids, hominins could certainly have dispersed from Eurasia into Africa. However, we are clear that the evidence is not overwhelming and that homoplasy may account for the hominin characters of *Graecopithecus*.

With less than thorough consideration of the details of our argument, Benoit and Thackeray[1] repeat the classic position points favouring an African origin of the hominins. An African origin of Hominini is well documented by the huge number of fossils that represent an unambiguous lineage of hominins from *Australopithecus* to *Homo,* probably also including *Ardipithecus, Orrorin* and *Sahelanthropus.* In addition, as argued by Huxley and Darwin, our nearest living relatives, *Pan* and *Gorilla,* are both exclusive to Africa, so it is most parsimonious to suggest that hominins arose there as well.

Neither argument is relevant to our conclusions. First, our results concern fossils that are about 3 million years older than the oldest *Australopithecus* and probably at least 600 000 years older than the oldest putative African hominin (*Sahelanthropus*). The completeness of the fossil record of hominins in Africa is not relevant to their origins, much as the fossil record of platyrrhines, which is exclusively American, is not relevant to their origins in Africa.

Second, dismissing the hypothesis of the presence of a late Miocene hominin in Europe ignores the large body of data demonstrating the widespread presence of hominids of modern aspect in Europe well before any appear in Africa. Nearly every phylogenetic analysis of Miocene apes, whether cladistic or not, concludes that dryopithecins and related taxa, which are exclusive to Europe, are hominids. This includes the analysis in Benoit and Thackeray[1]. The most comprehensive analyses have further concluded that they are hominines (African apes and humans) (e.g. Begun et al.[3,4]; Young and MacLatchy[5]). The same logic used to refute the Eastern Mediterranean origins hypothesis actually serves to support it. Because the overwhelming majority of Miocene hominid fossil taxa known are from Eurasia, it is reasonable to suppose that one or more of the extant subfamilies evolved there as well.

Benoit and Thackeray[1] state: 'Even if *Graecopithecus* can be attributed to Hominini, the fact that it is older than *Sahelanthropus* does not make it the basal-most representative of this clade.' We agree. We are well aware of the fact that geological age does not reveal phylogenetic position, nor does it establish divergence times. The age of the oldest known fossil of a particular taxon can only be interpreted as the currently known first appearance datum, not the origin, whether in time, geography or phylogeny. Our point is simply that the hypothesis that *Graecopithecus* is the oldest known hominin cannot be dismissed out of hand with the convenient invocation of homoplasy. The large body of data used to support this hypothesis must be addressed. Indeed, it is currently not possible to resolve the question of the most basal hominin. Neither *Sahelanthropus* nor *Graecopithecus* are known well enough to provide an unambiguous answer. Nonetheless, at 7.2 Ma, *Graecopithecus freybergi* remains the oldest candidate for this clade.

Benoit and Thackeray[1] also state: 'If *Graecopithecus* happens to be more derived than *Sahelanthropus*, then the evolutionary tree of Hominini would remain rooted in Africa and *Graecopithecus* would only represent an offshoot that dispersed out of Africa very early in the evolutionary history of hominins.'

This critique might have some relevance if we had actually reached the conclusion that *Graecopithecus* was derived relative to *Sahelanthropus*. However, we never stated that *Graecopithecus* was derived relative to *Sahelanthropus* in either P4 or canine root morphology. In fact, we state that there is variability in P4 root morphology in hominins, but that root fusion never occurs in Miocene apes and very rarely in *Pan*. We also never state, contra Benoit and

Thackeray[1], that canine root reduction is more derived in *Graecopithecus* than *Sahelanthropus.* We do not comment on the significance of this difference simply because it is obvious that the differences in canine root size between *Graecopithecus* and *Sahelanthropus* are well within ranges of within-sex variation in most hominoids – both fossil and extant. Once again, we are simply reporting that the canine root is reduced, as in hominins and to the distinction of other hominoids. Falsely attributing provocative conclusions to our work does not advance this debate.

Benoit and Thackeray[1] state: 'On the other hand, *Graecopithecus* might be closely related to *Ouranopithecus*, with which it has been synonymised for a long time or to other Eurasian apes, as suggested by previous cladistic analyses. In these cases, the evolutionary root of humankind would definitely remain in Africa.'

We provide extensive documentation of the differences between *Graecopithecus* and *Ouranopithecus*, none of which is addressed in Benoit and Thackeray[1]. We are not aware of any phylogenetic analysis that 'synonymises' these taxa, or either of them to other Eurasian apes. Only a sister clade relationship between *Ouranopithecus* and *Graecopithecus* would call our conclusions into question; however, we have presented ample evidence that this hypothesis of relationship is unlikely. We present evidence that *Graecopithecus* shares derived characters with hominins not found in *Ouranopithecus*. The most parsimonious interpretation of this distribution of characters is that *Ouranopithecus* predates the divergence of hominins and *Graecopithecus*. To assert that '*Graecopithecus* might be closely related to *Ouranopithecus*' without justification other than tradition, is not useful to this exchange.

About the cladistic analysis

Benoit and Thackeray[1] modified a character matrix published by Finarelli and Clyde[6], in turn modified from Begun et al.[4], to produce a cladistic analysis. Unfortunately, this character matrix is outdated in terms of both taxonomic units and character states, yielding a misleading and less parsimonious cladogram. More recent analyses (e.g. Begun et al.[3]; Young and MacLatchy[5]) yield very different cladograms from that in Benoit and Thackeray or Finarelli and Clyde but have the advantage of having been produced by researchers who compiled the character matrices from direct observation of the fossils, which is not the case for Finarelli and Clyde or Benoit and Thackeray. Finarelli and Clyde[6] were interested in the relationship between phylogeny and temporal sequence, which also influences their results. A revision of the original Begun et al.[4] data matrix, the most comprehensive published so far, is in preparation but is beyond the scope of this response.

It is unfortunate that when Benoit and Thackeray[1] state that 'none of the characters cited by Fuss et al. is strictly unique to Hominini, as thick enamel and megadonty' they omit the characters we actually say are unique to Hominini: reduction of premolar root complexity and canine root size. Nowhere in our publication do we say that thick enamel and megadonty are hominin synapomorphies. This mischaracterisation of our work only serves to polarise this debate.

It is clear that if *Graecopithecus* were found in Africa instead of Europe, its age and morphology would be taken as evidence that it is the earliest known hominin. *Chororapithecus* is accepted by many as an early gorilla, despite a very fragmentary sample and the fact that much more complete fossils with gorilla-like attributes are known from Europe.[7] But it is from Africa, where the earliest gorillas are supposed to be. The real problem is not morphology or preservation but a location that does not conform to the expectations of generations of palaeoanthropologists.

Our final quote from Benoit and Thackeray is most revealing of the confusion and artificial nature of this exchange: 'Our analysis supports the view that *Graecopithecus* is potentially an important taxon for the origin of Hominini, but this is not certain and deserves further investigation and more material.'

This is almost precisely the same conclusion we reached: 'Therefore, we submit that the dental root attributes of *Graecopithecus* suggest hominin affinities, such that its hominin status cannot be excluded. … More fossils are needed but at this point it seems likely that the Eastern Mediterranean needs to be considered as just as likely a place of hominine diversification and hominin origins as tropical Africa.'

Far from being an 'assertion that *Graecopithecus* belongs to Hominini' (Benoit and Thackeray[1(p.1)], we clearly present our results as preliminary and in need of additional fossils for confirmation. In contrast, Benoit and Thackeray[1] make numerous assertions about our results that are either unsubstantiated or inaccurate. The better way forward is through a thoughtful assessment of data and results as they are actually presented. We are hopeful that future contributions to this debate will maintain this focus.

References

1. Benoit J, Thackeray JF. A cladistic analysis of *Graecopithecus*. S Afr J Sci. 2017;113(11/12), Art. #a0238, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0238
2. Fuss J, Spassov N, Begun DR, Böhme M. Potential hominin affinities of *Graecopithecus* from the Late Miocene of Europe. PLoS ONE. 2017;12(5), e0177127, 23 pages. https://doi.org/10.1371/journal.pone.0177127
3. Begun DR, Nargolwalla MC, Kordos L. European Miocene hominids and the origin of the African ape and human clade. Evol Anthropol. 2012;21(1):10–23. https://doi.org/10.1002/evan.20329
4. Begun DR, Ward CV, Rose MD. Events in hominoid evolution. In: Begun DR, Ward CV, Rose MD, editors. Function, phylogeny, and fossils: Miocene hominoid evolution and adaptions. New York: Plenum Press; 1997. p. 389–415. https://doi.org/10.1007/978-1-4899-0075-3_18
5. Young NM, MacLatchy L. The phylogenetic position of *Morotopithecus*. J Hum Evol. 2004;46(2):163–184. https://doi.org/10.1016/j.jhevol.2003.11.002
6. Finarelli JA, Clyde WC. Reassessing hominoid phylogeny: Evaluating congruence in the morphological and temporal data. Paleobiology. 2004;30(4):614–651. https://doi.org/10.1666/0094-8373(2004)030<0614:RHPECI>2.0.CO;2
7. Begun DR. Fossil record of Miocene hominoids. In: Henke W, Tattersall I, editors. Handbook of paleoanthropology 2 – Primate evolution and human origins. Berlin: Springer; 2015. p. 1261–1332. https://doi.org/10.1007/978-3-642-39979-4_32

A temperature index in a Late Quaternary sequence at Wonderkrater, South Africa

AUTHOR:
J. Francis Thackeray[1]

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

CORRESPONDENCE TO:
Francis Thackeray

EMAIL:
Francis.thackeray@wits.ac.za

KEYWORDS:
Younger Dryas; age-depth relationships; pollen spectra; radiocarbon dates

HOW TO CITE:
Thackeray JF. A temperature index in a Late Quaternary sequence at Wonderkrater, South Africa. S Afr J Sci. 2018;114(5/6), Art. #a0268, 2 pages. http://dx.doi.org/10.17159/sajs.2018/a0268

PUBLISHED:
30 May 2018

Wonderkrater is a Late Quaternary archaeological site in the Limpopo Province of South Africa[1], and has a high-resolution Late Quaternary pollen sequence[2-4]. An initial multivariate analysis of pollen spectra facilitated the quantification of a temperature index (SSF1) for two borehole sequences.[5] An attempt to calibrate the SSF1 temperature indices was made by Thackeray[6]. Scott et al.[7] and Thackeray and Scott[8] identified the 'Younger Dryas' cooling event in Borehole 3. This palaeoclimatic phenomenon is generally known in the northern hemisphere, dated between 10 600 and 12 900 BP (calibrated years). In this study, palaeotemperature estimates from the initial multivariate study[5] are re-examined in the context of calibrated radiocarbon dates[7] in order to make comparisons with dates for the Younger Dryas in the northern hemisphere.

Age-depth relationships

On the basis of the SSF1 palaeotemperature index, Thackeray and Scott[8] recognised certain Wonderkrater samples as being associated with the Younger Dryas. Critical in this regard are dates for the relevant deposits. Scott et al.[3] and Scott[4] have used various techniques to resolve chronological issues. Whereas there is a linear relationship between age and depth for the last 16 000 years BP, the rate of deposition is not constant for earlier deposits for which the rate is slower and variable; hence the need for complex chronological modelling which presented a major challenge.[3,4] However, for purposes of this study, attention is focused on age-depth relationships for the Terminal Pleistocene and Holocene, within the last 16 000 years, for periods in which least squares linear regression is suitable. This offers a relatively simple but robust approach for dating postglacial deposits at Wonderkrater.

Three age-depth relationships were obtained by using calibrated radiocarbon dates (and associated standard deviations) for 12 Late Pleistocene and Holocene samples[7] from Borehole 3, using linear regression analyses. All 12 of these samples were selected by Scott et al.[7] because they were considered to be reliable. Others were deliberately excluded because they were outliers and were assumed to be associated with contamination by roots.

Equation 1 is obtained by relating mean values for calibrated radiocarbon dates (CAL-1, *y*-axis) and depth (D, *x*-axis), for the 12 samples selected by Scott et al.[7] in their Table 1:

$$\text{CAL-1} = 40.423\ D - 1431.30\ (r = 0.99) \qquad \text{Equation 1}$$

Equation 2 is obtained by relating depth to chronological values which are two standard deviations *below* the mean CAL-1 date for each sample:

$$\text{CAL-2} = 39.696\ D - 1487.98\ (r = 0.99) \qquad \text{Equation 2}$$

Equation 3 is obtained by relating depth to chronological values which are two standard deviations *above* the mean CAL-1 date for each sample:

$$\text{CAL-3} = 41.182\ D - 1413.32\ (r = 0.99) \qquad \text{Equation 3}$$

On the basis of depth, all three equations were applied in order to estimate dates for 50 samples from Borehole 3. Estimates for CAL-1 relative to the SSF1 palaeotemperature index are shown in Figure 1.

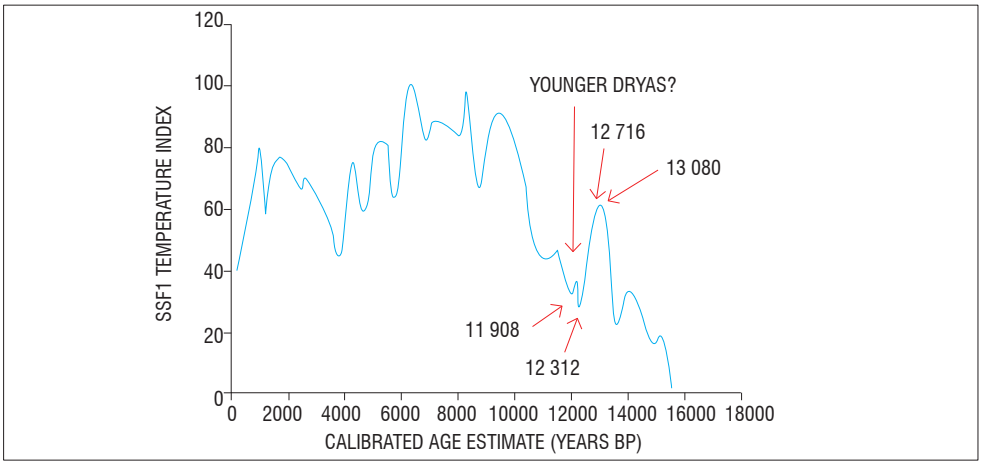


Figure 1: A temperature index (SSF1) based on multivariate analysis of pollen spectra from Wonderkrater in relation to estimated CAL-1 dates (cal years BP) based on Equation 1. A cooling event appears to occur after 12 716 cal years BP – here associated with the onset of the Younger Dryas which is well documented in the northern hemisphere where it is dated between 12 900 and 11 600 cal years BP.

Dates in relation to temperature index

Notable in Figure 1 and Table 1 is that relatively low temperatures are identified at 12 312 cal years BP (sample 5609), continuing to 12 100 cal years BP (sample 5606) and further to 11 908 cal years BP (sample 5605). This cool event appears to be initiated at 12 716 cal years BP (sample 5611) after a slightly warmer period at 13 080 cal years BP (sample 5613).

Table 1: Wonderkrater samples listed by number; depth (cm); calibrated dates (CAL-1, CAL-2 and CAL-3 estimates based on Equations 1–3); and SSF1 temperature indices based on multivariate analyses of pollen spectra from Borehole 3, Wonderkrater sequence[5]. CAL-2 and CAL-3 are the lower and upper limits (two standard deviations), respectively, for estimates of the mean calibrated dates (CAL-1).

Sample #	Depth	CAL-2 (lower limit)	CAL-1 (mean)	CAL-3 (upper limit)	SSF1
5599	300	10 420	10 695	10 941	46
5601	310	10 817	11 099	11 353	43
5603	320	11 214	11 504	11 765	46
5605	330	11 611	11 908	12 177	32
5606	335	11 810	12 110	12 382	37
5609	340	12 008	12 312	12 588	29
5611	350	12 405	12 716	13 000	57
5613	359	12 762	13 080	13 371	61
5615	365	13 001	13 323	13 618	37

In the context of lower (CAL-2) and upper (CAL-3) limits for estimated CAL-1 dates (Table 1), the results from this study are consistent with calibrated radiocarbon dates for the Younger Dryas in the northern hemisphere. At Wonderkrater, sample 5611 at 12 716 cal years BP is close to the date of 12 900 cal years BP for the onset of the Younger Dryas in the northern hemisphere.[9]

The Younger Dryas in Borehole 4

The Younger Dryas is also represented in Borehole 4. On the basis of palaeotemperature indices for this sequence, a cooling episode can be identified from at least one sample.[3,4] The calibrated date of circa 12 200 BP for this sample from Borehole 4 is almost identical to the mean date (12 433 cal BP) for the three Younger Dryas samples from Borehole 3.

The Younger Dryas and a cosmic impact?

Among others, Kennett et al.[10] have claimed that the Younger Dryas cooling event may be associated with a cosmic impact of some kind. Evidence given in support of this possible cosmic impact, at least for the northern hemisphere, are spikes in nanodiamonds[10,11], platinum[12], magnetic and glassy impact-related spherules, high-temperature minerals and melt glass, carbon spherules and/or osmium at the onset of the Younger Dryas[9].

A spike in nanodiamonds has been reported for Younger Dryas deposits in Mexico.[13] As yet, no corresponding site has been reported for the southern hemisphere. In the context of dates given here for palaeotemperature indices for the Wonderkrater sequence, it is recommended that exploratory analyses be undertaken on specific samples (notably 5605, 5606 and 5609 from Borehole 3) from this important South African site, to test whether or not there is evidence for nanodiamonds or other indicators of a cosmic impact that may have affected climates globally sometime after 12 900 cal years BP.

Acknowledgements

This work was supported by the National Research Foundation (South Africa) and the DST/NRF Centre of Excellence for the Palaeosciences. I thank Louis Scott for his comments.

References

1. McCarthy TS, Ellery WN, Backwell L, Marren P, De Klerk B, Tooth S, et al. The character, origin and palaeoenvironmental significance of the Wonderkrater spring mound, South Africa. J Afr Earth Sci. 2010;58(1):115–126. http://dx.doi.org/10.1016/j.jafrearsci.2010.02.004
2. Scott L. A Late Quaternary pollen record from the Transvaal bushveld, South Africa. Quat Res. 1982;17:339–370. https://doi.org/10.1016/0033-5894(82)90028-X
3. Scott L, Neumann FH, Brook GA, Bousman CB, Norström E, Metwally A. Terrestrial fossil-pollen evidence of climate change during the last 26 thousand years in southern Africa. Quat Sci Rev. 2012;32:100–118. https://doi.org/10.1016/j.quascirev.2011.11.010
4. Scott L. Fluctuations of vegetation and climate over the last 75 000 years in the Savanna Biome, South Africa: Tswaing Crater and Wonderkrater pollen sequences reviewed. Quat Sci Rev. 2016;145:117–133. https://doi.org/10.1016/j.quascirev.2016.05.035
5. Scott L, Thackeray JF. Multivariate analysis of late Pleistocene and Holocene pollen spectra from Wonderkrater, Transvaal, South Africa. S Afr J Sci. 1987;83:93–98.
6. Thackeray JF. Calibration of temperature indices from a Late Quaternary terrestrial sequence at Wonderkrater, South Africa. Quat Int. 1999;57/58:225–227. https://doi.org/10.1016/S1040-6182(98)00063-9
7. Scott L, Holmgren K, Talma AS, Woodborne S, Vogel JC. Age interpretation of the Wonderkrater spring sediments and vegetation change in the Savanna Biome, Limpopo Province, South Africa. S Afr J Sci. 2003;99:484–488.
8. Thackeray JF, Scott L. The Younger Dryas in the Wonderkrater sequence, South Africa? Ann Transv Mus. 2006;43:111–112.
9. Kennett JP, Kennett DJ, Culleton BJ, Tortosa JEA, Bunch TE, Erlandson JM, et al. Reply to Holliday and Boslough et al: Synchroneity of widespread Bayesian-modeled ages support Younger Dryas impact hypothesis. Proc Natl Acad Sci USA. 2015;112(49):E6723–E6724. https://doi.org/10.1073/pnas.1520411112
10. Kennett DJ, Kennett JP, West A, Mercer C, Que Hee SS, Bement L, et al. Nanodiamonds in the Younger Dryas boundary sediment layer. Science. 2009;323(5910):94. https://doi.org/10.1126/science.1162819
11. Kurbatov AV, Mayewski P, Steffensen J, West A, Kennett D, Kennett J, et al. Discovery of a nanodiamond-rich layer in the Greenland ice sheet. J Glaciol. 2010;56(199):747–757. https://doi.org/10.3189/002214310794457191
12. Petaev MI, Huang S, Jacobsen SB, Zindler A. Large Pt anomaly in the Greenland ice core points to a cataclysm at the onset of Younger Dryas. Proc Natl Acad Sci USA. 2013;110(32):12917–12920. https://doi.org/10.1073/pnas.1303924110
13. Israde-Alcantara I, Bischoff JL, Domínguez-Vázquez G, Li H-C, DeCarli PS, Bunch TE, et al. Evidence from Central Mexico supporting the Younger Dryas extraterrestrial impact hypothesis. Proc Natl Acad Sci USA. 2012;109(13):E738–E747. https://doi.org/10.1073/pnas.1110614109

Psychrophiles: Ecological significance and potential industrial application

AUTHOR:
Amira Hamdan[1]

AFFILIATION:
[1]Oceanography Department, Alexandria University, Alexandria, Egypt

CORRESPONDENCE TO:
Amira Hamdan

EMAIL:
amira_hamdan1978@yahoo.com

DATES:
Received: 31 July 2017
Revised: 15 Dec. 2017
Accepted: 17 Jan. 2018
Published: 30 May 2018

KEYWORDS:
frozen ecosystem; biodiversity; cold-active proteins; carotenoid pigments; food industry; pharmaceutical industry

HOW TO CITE:
Hamdan A. Psychrophiles: Ecological significance and potential industrial application. S Afr J Sci. 2018;114(5/6), Art. #2017-0254, 6 pages. http://dx.doi.org/10.17159/sajs.2018/20170254

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

The bulk of earth's biosphere is cold (<5 °C) that sustains a broad diversity of microbial life by triggering physiological response(s) to ensure survival in cold and frozen ecosystems. The strategy of adaptation to cold environments includes changes in membrane composition and induction of a set of specific cold-active proteins, polyunsaturated lipids and exopolysaccharides. These adaptive features provide an enormous natural reservoir of enzymes that function effectively in cold environments, and these cold-active enzymes have been targeted for innovative applications useful to humankind. This review provides an overview of the existence, distribution and adaptation strategies of psychrophilic microorganisms worldwide with great emphasis on their recently emerged industrial applications in the textile industry, food and dairy industry, brewing and wine industry, laundry detergent industry, and others.

Significance:

- The outcome of these studies may also help in the exploration of the possibility of life in distant frozen planets.

Introduction

A great proportion of the earth's biosphere (>85%) permanently experiences temperatures below 5 °C.[1] The largest coverage of these cold environments is successfully colonised by a wide diversity of extremophilic microorganisms, including bacteria, archaea, yeasts, filamentous fungi and algae.[2,3] In 1887, Forster first reported the existence of bacteria capable of growth at 0 °C, especially in sea water and ocean fish.[4] Cold-adapted or cold-loving microbes are termed 'psychrophiles', and have cardinal growth temperatures (minimum, optimum and maximum) at or below 0 °C, 15 °C and 20 °C, respectively, while microorganisms that withstand cold temperatures with a higher growth optimum and maximum (above 25 °C) are called 'psychrotolerant'.[5] The ability to thrive at such low temperatures requires a vast array of adaptations to maintain the metabolic rates and sustain growth compatible with life in these severe environmental conditions.[6] Because of their attractive properties, extensive research on psychrophilic microorganisms has been conducted to understand their survival strategies that include genetic and acclimation processes and adaptation mechanisms.[7]

The purpose of this review is to summarise the tremendous significance of this group of microorganisms in fundamental research, pharmaceuticals, medicine and recent biotechnological applications.

Psychrophiles: Historical background

In 1887, Forster early investigated the isolation of microorganisms from fish preserved by cold temperatures that were bioluminescent.[8] Later in 1892, Forster reported other bacterial isolates from various environments (natural water, food, surface and intestines of freshwater and seawater fish) that were able to grow at 0 °C.[4] Schmidt-Nielsen (1902) first mentioned the term 'psychrophile' to describe bacteria capable of growing at 0 °C.[9] However, Müller (1903) criticised this term and demonstrated that while these organisms were able to grow at low temperature they actually grew more rapidly at elevated temperatures.[10] Thus, Kruse (1910) suggested that these bacteria might be better called 'psychrotolerant', but this suggestion or others of similar nature did not gain acceptance and the term 'psychrophile' has retained.[11] This confusion was ended in 1975, when Morita proposed a definition to the term 'psychrophile' for cold-adapted or cold-loving microbes, having minimum, optimum and maximum growth temperatures at or below 0 °C, 15 °C and 20 °C, respectively.[12]

Habitats and biodiversity

Cold habitats dominate the vast majority of our planet, covering three-quarters of the earth's surface, and span from the Arctic to the Antarctic and from high-mountain regions to the deep ocean.[13,14] The major fraction of this low temperature environment is represented by the deep sea (90% of the ocean volume), followed by snow (35% of land surface), permafrost (24% of land surface), sea ice (13% of the earth's surface) and finally glaciers (10% of land surface). Other cold environments are cold-water lakes, cold soils, cold deserts and caves.[15] These earth dominant environments are successfully colonised by enormously diverse communities of psychrophilic bacteria, archaea, algae, yeast[16-20], insects[21] and fish[22], that are able to thrive and even maintain metabolic activity at subzero temperatures.

Bacteria represent very important members of the sea ice habitat, including many unique taxa.[23] Heterotrophic gas-vacuolate bacteria, not reported in other marine habitats, have been discovered in and near sea ice.[24] Among those cold-adapted bacteria, the genus *Colwellia* provides an unusual case. Members of this genus produce extracellular enzymes capable of degrading high-molecular-weight organic compounds. These traits make *Colwellia* species important to carbon and nutrient cycling wherever they occur in the cold marine environment, from contaminated sediments to ice formations as analogs for possible habitats on other planets and moons (e.g. Mars and Europa).[13]

Representatives of the family Vibrionaceae are among the most commonly reported bacteria to populate almost all extreme environments.[25] Nevertheless, a wide range of phylogenetic diversity within the genera

Achromobacteria, Alcaligenes, Altermonas, Aquaspirillum, Arthrobacter, Bacillus, Bacteroides, Brevibacterium, Clostridium, Colwellia, Cytophaga, Flavobacterium, Gelidibacter, Methanococcoides, Methanogenium, Methanosarcina, Microbacterium, Micrococcus, Moritella, Octadecabacter, Phormidium, Photobacterium, Polaribacter, Polaromonas, Pseudomonas, Psychroserpens, Shewanella, *Psychrobacter* and *Vibrio* have been found to be psychrophilic across the domain Bacteria.[26-29]

In general, fungi are relatively rare in deep sea habitats compared to bacteria.[30] Fungal isolates reported in frozen environments belong mainly to the genera *Rhodotorula, Penicillium, Ustilago*, *Alternaria, Aureobasidium, Cladosporium, Geomyces, Ulocladium, Valsa* and *Verticillium*.[31-33]

Existence and metabolic adaptation

Survival of psychrophiles in harsh and extremely cold environments is interesting and requires a vast array of unique adaptive features from all their cellular components.[34,35] Chintalapati et al.[36] categorised three phases of cold-shock response. Phase I, Acclimation Phase, immediately follows cold exposure, leading to reduced growth rate as a result of reduced membrane fluidity, and several cold-shock proteins are produced. During Phase II, Recovery Phase, cells are considered 'cold-adapted' and resume growth and bulk protein synthesis restarts. During Phase III, Stationary Phase, non-cold-shock proteins are synthesised, allowing cells to proceed to slow-rate growth at low temperature.[37]

In general, bacteria owe the ability to cope with such challenges to complex strategies. One important strategy is directed towards their extremely efficient DNA repair mechanism sustained under frozen conditions.[38] This evidence for active DNA repair mechanism in icy environments is a true reflection of their physiological potential and survival in frozen substrates for extended time frames of up to 600 000 years.[39]

Regulation of membrane fluidity

Another important and the most frequent adaptive strategy relates to the ability of the cell to regulate or modulate the fluidity of the membrane in freezing environments. The membrane is the first barrier that can sense environmental changes and it acts as an interface between external and internal environments, so as to overcome the deleterious effects of harsh conditions.[36] Shivaji and Prakash[35] reported that membranes become more rigid at cold temperatures, which activates a membrane-associated sensor resulting in upregulation of genes involved in membrane fluidity modulation for exchange of metabolites from and to the cell.

Consequently, changes in the membrane lipid composition facilitate this process. This is achieved by modifications in the lipids' fatty acyl chains that serve to maintain optimum membrane fluidity.[40] In general, lower growth temperatures govern the activation of a group of cold-shock-activated enzymes called 'desaturases' which convert saturated acyl fatty acids to unsaturated ones and aid the increase in the proportion of unsaturated acyl chains, reduction in the acyl chain length and increased methyl-branched fatty acids.[36] Fungal cell membranes show evidence of similar changes at cold ecosystems to maintain their fluid state. The degree of unsaturated fatty acids increases at low temperatures in *Candida, Leucosporidium* and *Torulopsis* as reported by Kerekes and Nagy[41].

Carotenoid pigments

Predominance of carotenoid pigments, in several bacteria isolated from Antarctic sea ice, has been reported to play an important role in the maintenance of membrane structure, fluidity and protection from UV radiation.[42] Accumulation of C-50 carotenoid, Bacterioruberin, observed in the psychrotrophic strain, *Arthrobacter agilis*, was postulated to play a crucial role in the regulation of membrane fluidity at low temperatures.[43] Furthermore, several polar and non-polar carotenoid pigments, synthesised by *Micrococcus roseus* and *Sphingobacterium antarcticus* Antarctic strains, were found to bind vesicles, made of both synthetic and natural lipids, and to rigidify them.[44]

Anti-freeze proteins

Anti-freeze proteins (AFPs) are ice-binding proteins that have the ability to decrease the freezing point of water and show extracellular ice re-crystallisation inhibition activity during latter stages of the warming cycle.[45,46] Furthermore, AFPs have been well known in promoting super cooling of the body fluids at subzero temperatures to prevent freezing of blood in polar fish.[47] In addition, those proteins have been reported in insects and plants.[48,49] Duman and Olsen[50] first demonstrated the presence of AFPs in cold-adapted bacteria and the bacterium *Moraxella* sp. was the first reported Antarctic strain to produce an AFP.[51] Some psychrophilic AFPs have been purified from cell extracts of *Micrococcus cryophilus, Pseudomonas putida* and *Rhodococcus erythropolis*.[52]

Cryoprotectants

Cryoprotectants are exopolymeric substances (e.g. sugars, alcohols and amino acids), produced in high amounts and believed to prevent cold-induced aggregation of proteins and maintain optimum membrane fluidity under unfavorable low temperature.[53,54] In 1994, Ko et al.[55] demonstrated the growth-enhancing effect of glycine betaine on *Listeria monocytogenes* at low temperature. Furthermore, trehalose showed evidence in preventing protein denaturation and aggregation in psychrophilic bacteria.[56] In fungi, trehalose is an important stress protectant and stabiliser of membranes, accumulated in fungal hyphae in large quantities at low temperatures.[57]

Cold-shock proteins

Cold-shock proteins have been extensively characterised and are considered the most prominent response of cells to cold shock in order to counteract the detrimental effect of temperature downshift; they play a critical role in cold adaptation.[58] Cold-shock proteins have been functionally linked to the regulation of cellular protein synthesis, particularly at the level of transcription and initiation of translation. They also act as chaperones by preventing the formation of mRNA secondary structures (mRNA 'folding') and maintenance of chromosome structure.[59,60] These stress proteins, together with cold-acclimation proteins, have been detected and exclusively overexpressed during the entire growth of several Antarctic bacteria.[56] Furthermore, Kawahara et al.[61] believed that the cold acclimation protein (Hsc 25) produced in an ice-nucleating bacterium *Pantoea ananas* KUIN-3, was capable of refolding cold-denatured enzymes that sustain biological activities following an abrupt temperature downshift.

Cold-active enzymes

Cold-active or 'cold-adapted' enzymes are those enzymes that display high catalytic efficiency at low temperature compared to their mesophilic counterparts.[62] Furthermore, Struvay and Feller[63] reported multiple adaptive features developed by constrained psychrophiles to design enzymes perfectly compatible to the given environment. In the first attempt, psychrophiles produce enzymes that are highly flexible in structure, having an up to 10-fold higher specific activity (k_{cat}) than their mesophilic homologues, thus providing better access to the active site of substrates at lower temperatures.[64] These cold-active enzymes offset the inhibitory effect of low temperatures on reaction rates and maintain adequate metabolic fluxes to the growing organism. Secondly, the apparent maximal activity of cold-active enzymes is shifted towards low temperatures to cope with the increased viscosity of the aqueous environment induced under cold conditions. In this context, these enzymes are heat labile and frequently inactivated at temperatures that are not detrimental for their mesophilic counterparts.[65] Finally, apart from their high catalytic efficiency, the adaptation to cold is not usually perfect, as the specific activity exhibited by most psychrophilic enzymes around 0 °C remains generally lower than that of their mesophilic counterparts at 37 °C.

Pioneering studies compared cold-active enzymes to their conventional mesophilic forms and observed significant differences in amino acid composition, mostly in the active site domain, responsible for their activity under cold conditions.[66] The criteria for structural alterations include increased clusters of glycine residues (providing local mobility)

and reduction in proline and arginine residues in loops, which provides enhanced chain flexibility between secondary structures and the capability of forming multiple salt bridges and hydrogen bonds, respectively.[6] However, surprisingly, Aghajari et al.[67] reported great similarity in the three-dimensional structure of a cold-adapted α-amylase enzyme of a psychrophilic bacterium, *Alteromonas hulopfuncris*, to those of mammalian α-amylases.

A variety of cold-active enzymes – xylanases and laminarases[68], chitinases[69], α-amylase and β-galactosidase[70,71], lipases and proteases[20,72], aminopeptidase[73] and protein-tyrosine phosphatase[74] – have been reported in cold-adapted bacteria to aid in their survival under cold conditions.

Biotechnological applications

Over recent years, considerable attention has been focused on psychrophilic microorganisms that proliferate in extremely cold niches and the fascinating biotechnological potential of their cold adaptation. Although Witter[11] reported the significant opportunity of psychrophilic bacteria in the dairy industry and in keeping milk fresh under refrigeration facilities for longer holding times as far back as 1961, studies on psychrophiles have been accentuated by the advancements and applications in avoiding potentially disastrous situations in various industries, including those concerned with food production, waste processing, mining, environmental bioremediations, agriculture, and medicine and molecular diagnostics.[75,76]

Bioremediation

Because of increasing human activities and demand for fossil fuel energy, spillage of petroleum products is a growing menace to the environment. Psychrophiles hold tremendous potential as 'environmental cleaners' to successfully degrade pollutants of petroleum hydrocarbons in extreme cold conditions.[77,78] Whyte et al.[79] reported that the microbial catabolic pathways responsible for the degradation of petroleum hydrocarbons, including *n*-alkanes and polycyclic aromatic hydrocarbons, are widespread in cold regions. Furthermore, several indigenous, cold-adapted microbial populations that use petroleum hydrocarbons have been detected, including *Rhodococcus*[80], *Pseudomonas*[81] and *Pseudoalteromomas*[82].

Food industry

It has indeed emerged that cold-active enzymes represent an extremely powerful tool in the food industry. There is an increasing industrial trend to treat foodstuffs under low temperature conditions in order to avoid detrimental effects on taste, texture and nutritional value, and also to save energy. Cold-active β-galactosidase, which hydrolyses lactose to glucose and galactose at refrigerating temperature, is a potentially important enzyme in the dairy industry. It can be used to produce lactose-free milk-derived foods for lactose-intolerant people, who represent approximately 30% of the world population.[83] Furthermore, it can be used to convert lactose in whey to D-tagatose – a natural high-added-value sweetener with low caloric and glycemic index.[63] Hoyoux et al.[84] patented a cold-active β-galactosidase from an Antarctic psychrophile, *Pseudoalteromonas haloplanktis*, for its capacity to hydrolyse lactose during milk storage at low temperatures.

In this context, application of cold-active pectinases in the fruit juice industry is highly demanding to retain the quality and nutritional properties of the fruit juice and facilitate different processing steps – liquefaction, clarification and juice extraction – at low temperatures. Furthermore, residual enzyme activity could be easily eliminated by enzyme inactivation by moderate heat input after treatment.[85] In wine industries, cold-active pectinases, isolated from yeasts and fungi, are believed to increase the production and retention of volatile compounds, thereby improving the aromatic profile of wines.[86]

Also, it is worth mentioning that cold-active xylanases, hydrolysing β-1,4-xylan present in all flours, are one of the key ingredients of industrial dough conditioners used at cool temperatures required for dough resting to improve bread quality. Following careful baking trials, xylanase from the Antarctic bacterium *Pseudoalteromonas haloplanktis*, effectively improved the mechanical dough properties and final bread quality with a positive effect on loaf volume.[63]

Medical and pharmaceutical applications

There is a growing interest in studying psychrophilic bacteria as new tools in pharmaceutical and cosmetic applications. Consequently, several promising psychrophilic strains were detected as a valuable source of new active antimicrobial compounds at low temperatures, as reported by Tomova et al.[87] Recently, compounds produced by the halophilic Antarctic actinomycete *Nocardioides* sp. strain A-1, with antimicrobial activity against *Xanthomonas oryzae* which causes bacterial blight disease in rice-producing countries, exhibited promising application in agriculture for plant protection.[88] In addition, it was found that Antarticine-NF3, an antifreeze glycoprotein produced by the Antarctic bacterium *Pseudoalteromonas*, is effective for scar treatment and has been included in cosmetic regeneration creams.[89]

Furthermore, polyunsaturated fatty acids, produced to improve membrane permeability and nutrient transport in psychrophilic bacteria, may constitute an economic alimentary source for aquaculture industries with favorable activities on cholesterol and triglyceride transport for optimal nervous system and cardiovascular health.[90]

At the industrial level, the polar yeast *Candida antarctica* is used to produce two cold-active lipases, A and B, sold as Novozym 435 by Novozymes (Bagsvaerd, Denmark) involved in a very large number of applications related to pharmaceuticals and cosmetics.[91]

Detergent and fabric industry

Considering the versatile properties of cold-active enzymes – such as high catalytic efficiencies at low temperature, lower thermal stability and novel substrate specificities – they offer a large reservoir of potentially novel biotechnological exploitation.[75] Nowadays, the detergent field, which represents 30–40% of all enzymes produced worldwide, requires more enzymes that are capable of working at low temperatures in the context of energy saving. Cold-adapted enzymes actively used in detergent formulations – lipases, proteases and α-amylases – are systematically introduced to improve the efficiency of detergents and reduce the amount of chemicals used in order to protect the texture and colours of fabrics and reduce wear and tear during washing.[6,64,92] Currently, cold-active subtilisins, isolated from Antarctic *Bacillus* species, are incorporated in cold-active detergents that combine storage alkaline stability and cold activity required for optimal washing results.[63]

Furthermore, the application of cold-adapted cellulases in fabric production and denim finishing could increase the smoothness and softness of tissues, decolourisation of textile effluents and textile bleaching and allows the development of environmentally friendly technologies in fibre processing.[93]

Conclusion

Over recent years, a wealth of knowledge has been accumulated on psychrophilic microorganisms and their cold-shock response catalysts. Because of their unique biological features, psychrophiles can be significantly exploited in biotechnological industries such as the pharmaceutical, enzyme production, bioremediation, biosensor, cosmetic, agriculture, domestic purposes and textile industries. The outcome of these studies may also help to explore the possibility of life in distant frozen planets and their satellites.

References

1. Gounot AM. Microbial life in permanently cold soils. In: Margesin R, Schinner F, editors. Cold-adapted organisms, ecology, physiology, enzymology and molecular biology. New York: Springer; 1999. p. 4–15. https://doi.org/10.1007/978-3-662-06285-2_1
2. Kirchman DL, Moran XA, Ducklow H. Microbial growth in the polar oceans – role of temperature and potential impact of climate change. Nat Rev Microbiol. 2009;7:451–459. https://doi.org/10.1038/nrmicro2115

3. Laybourn-Parry J, Pearce D. Heterotrophic bacteria in Antarctic lacustrine and glacial environments. Polar Biol. 2016;39:2207–2225. https://doi.org/10.1007/s00300-016-2011-1

4. Morita R. Psychrophilic bacteria. Bacteriol Rev. 1975;39:144–167.

5. Helmke E, Weyland H. Psychrophilic *versus* psychrotolerant bacteria. Occurrence and significance in polar and temperate marine habitats. Cell Mol Biol. 2004;50:553–561.

6. Feller G, Gerday C. Psychrophilic enzymes: Hot topics in cold adaptation. Nat Rev Microbiol. 2003;1:200–208. https://doi.org/10.1038/nrmicro773

7. Zakhia F, Jungblut A-D, Taton A, Vincent W, Wilmotte A. Cyanobacteria in cold ecosystems. In: Margesin R, Schinner F, Marx J-C, Gerday C, editors. Psychrophiles: From biodiversity to biotechnology. Berlin: Springer; 2008. p. 121–135. https://doi.org/10.1007/978-3-540-74335-4_8

8. Georlette D, Blaise V, Collins T, D'Amico S, Gratia E, Hoyoux A, et al. Some like it cold: Biocatalysis at low temperatures. FEMS Microbiol Rev. 2004;28:25–42. https://doi.org/10.1016/j.femsre.2003.07.003

9. Feller G, Gerday C. Psychrophilic enzymes: Molecular basis of cold adaptation. Cell Mol Life Sci. 1997;53:830–841. https://doi.org/10.1007/s000180050103

10. Morita R, Moyer C. Psychrophiles, origin of. Encyclopedia of biodiversity. 2001;4:917–924. https://doi.org/10.1016/B0-12-226865-2/00362-X

11. Witter LD. Psychrophilic bacteria – A review. J Dairy Science. 1961;44:983–1015. https://doi.org/10.3168/jds.S0022-0302(61)89851-2

12. Junge K, Christner B, Staley J. Diversity of psychrophilic bacteria from sea ice – and glacial ice communities. In: Horikoshi K, editor. Extremophiles handbook. Tokyo: Springer; 2011. p. 793–815.

13. Deming JW, Eicken H. Life in ice. In: Sullivan WT, Baross JA, editors. Planets and life: The emerging science of astrobiology. Cambridge: Cambridge University Press; 2007. p. 292–312.

14. Rodrigues DF, Tiedje JM. Coping with our cold planet. Appl Environ Microbiol. 2008;74:1677–1686. https://doi.org/10.1128/AEM.02000-07

15. Lauro FM, Bartlett DH. Prokaryotic lifestyles in deep sea habitats. Extremophiles. 2008;12:15–25. https://doi.org/10.1007/s00792-006-0059-5

16. Cavicchioli R. Cold-adapted Archaea. Nat Rev Microbiol. 2006;4:331–343. https://doi.org/10.1038/nrmicro1390

17. Kalanetra KM, Bano N, Hollibaugh T. Ammonia-oxidizing Archaea in the Arctic ocean and Antarctic coastal waters. Environ Microbiol. 2009;11:2434–2445. https://doi.org/10.1111/j.1462-2920.2009.01974.x

18. Margesin R, Miteva V. Diversity and ecology of psychrophilic microorganisms. Res Microbiol. 2011;162:346–361. https://doi.org/10.1016/j.resmic.2010.12.004

19. Buzzini P, Branda E, Goretti M, Turchetti B. Psychrophilic yeasts from worldwide glacial habitats: Diversity, adaptation strategies and biotechnological potential. FEMS Microbiol Ecol. 2012;82:217–241. https://doi.org/10.1111/j.1574-6941.2012.01348.x

20. Lamilla C, Pavez M, Santos A, Hermosilla A, Llanquinao V, Barrientos L. Bioprospecting for extracellular enzymes from culturable Actinobacteria from the South Shetland Islands, Antarctica. Polar Biol. 2017;40:719–726. https://doi.org/10.1007/s00300-016-1977-z

21. Doucet D, Walker VK, Qin W. The bugs that came in from the cold: Molecular adaptations to low temperatures in insects. Cell Mol Life Sci. 2009;66:1404–1418. https://doi.org/10.1007/s00018-009-8320-6

22. Giordano D, Russo R, Di Prisco G, Verde C. Molecular adaptations in Antarctic fish and marine microorganisms. Mar Genomics. 2012;6:1–6. https://doi.org/10.1016/j.margen.2011.09.003

23. Helmke E, Weyland H. Bacteria in sea ice and underlying water of the eastern Weddell Sea in midwinter. Mar Ecol Prog Ser. 1995;117:269–287. https://doi.org/10.3354/meps117269

24. Staley JT, Irgens RL, Herwig RP. Gas vacuolated bacteria found in Antarctic sea with ice algae. Appl Environ Microbiol. 1989;55:1033–1036.

25. D'Amico S, Collins T, Marx JC, Feller G, Gerday C. Psychrophilic microorganisms: Challenges for life. EMBO Rep. 2006;7:385–389. https://doi.org/10.1038/sj.embor.7400662

26. Staley JT, Gosink JJ. Poles apart: Biodiversity and biogeography of sea ice bacteria. Annu Rev Microbiol. 1999;53:189–215. https://doi.org/10.1146/annurev.micro.53.1.189

27. Deming JW. Psychrophiles and polar regions. Curr Opin Microbiol. 2002;5:301–309. https://doi.org/10.1016/S1369-5274(02)00329-6

28. Miteva V. Bacteria in snow and glacier ice. In: Margesin R, Schinner F, Marx J-C, Gerday C, editors. Psychrophiles: From biodiversity to biotechnology. Berlin: Springer; 2008. p. 31–50. https://doi.org/10.1007/978-3-540-74335-4_3

29. Zhang X, Ma X, Wang N, Yao T. New subgroup of *Bacteroidetes* and diverse microorganisms in Tibetan plateau glacial ice provide a biological record of environmental conditions. FEMS Microbiol Ecol. 2008;67:21–29. https://doi.org/10.1111/j.1574-6941.2008.00604.x

30. Bass D, Howe A, Brown N, Barton H, Demidova M, Michelle H, et al. Yeast forms dominate fungal diversity in the deep oceans. Proc Roy Soc B Biol Sci. 2007;274:3069–3077. https://doi.org/10.1098/rspb.2007.1067

31. Robinson CH. Cold adaptation in Arctic and Antarctic fungi. New Phytologist. 2001;151:341–353. https://doi.org/10.1046/j.1469-8137.2001.00177.x

32. Ozerskaya SM, Ivanushkina NE, Kochkina GA, Fattakhova RN, Gilichinsky DA. Mycelial fungi in cryopegs. Int J Astrobiol. 2004;4:327–331. https://doi.org/10.1017/S1473550405002260

33. D'Elia T, Veerapaneni R, Theraisnathan V, Rogers SO. Isolation of fungi from Lake Vostok accretion ice. Mycologia. 2009;101:751–763. https://doi.org/10.3852/08-184

34. Bej AK, Aislabie J, Atlas RM. Polar microbiology: The ecology, biodiversity and bioremediation potential of microorganisms in extremely cold environments. Boca Raton, FL: CRC Press; 2010.

35. Shivaji S, Prakash J. How do bacteria sense and respond to low temperature? Arch Microbiol. 2010;192:85–95. https://doi.org/10.1007/s00203-009-0539-y

36. Chintalapati S, Kiran MD, Shivaji S. Role of membrane lipid fatty acids in cold adaptation. Cell Mol Biol. 2004;50:631–642.

37. Weber MH, Marahiel MA. Bacterial cold shock responses. Sci Progr. 2003;86:9–75. https://doi.org/10.3184/003685003783238707

38. Christner BC. Incorporation of DNA and protein precursors into macromolecules by bacteria at -15°C. Appl Environ Microbiol. 2002;68:6435–6438. https://doi.org/10.1128/AEM.68.12.6435-6438.2002

39. Johnson SS, Hebsgaard MB, Christensen TR, Mastepanov M, Nielsen R, Munch K, et al. Ancient bacteria show evidence of DNA repair. Proc Natl Acad Sci USA. 2007;36:14401–14405. https://doi.org/10.1073/pnas.0706787104

40. Russell NJ. Membrane components and cold sensing. In: Margesin R, Schinner F, Marx J-C, Gerday C, editors. Psychrophiles: From biodiversity to biotechnology. Berlin: Springer; 2008. p. 177–190. https://doi.org/10.1007/978-3-540-74335-4_11

41. Kerekes R, Nagy G. Membrane lipid composition of a mesophilic and psychrophilic yeast. Acta Alimentaria. 1980;9:93–98.

42. Armstrong GA. Genetics of eubacterial carotenoid biosynthesis: A colourful tale. Annu Rev Microbiol. 1997;51:629–659. https://doi.org/10.1146/annurev.micro.51.1.629

43. Fong NJC, Burgess ML, Barrow KD, Glenn DR. Carotenoid accumulation in the psychrotrophic bacterium *Arthrobacter agilis* in response to thermal and salt stress. Appl Microbiol Biotechnol. 2001;56:750–756. https://doi.org/10.1007/s002530100739

44. Chattopadhyay MK, Jagannadham MV. Maintenance of membrane fluidity in Antarctic bacteria. Polar Biol. 2001;24:386–388. https://doi.org/10.1007/s003000100232

45. Gilbert JA, Hill PJ, Dodd CE, Laybourn-Parry J. Demonstration of antifreeze protein activity in Antarctic lake bacteria. Microbiology. 2004;150:171–180. https://doi.org/10.1099/mic.0.26610-0

46. Christner BC. Bioprospecting for microbial products that affect ice crystal formation and growth. Appl Microbiol Biotechnol. 2010;85:481–489. https://doi.org/10.1007/s00253-009-2291-2

47. Jia Z, DeLuca CI, Chao H, Davies PL. Structural basis for the binding of a globular antifreeze protein to ice. Nature. 1996;384:285–288. https://doi.org/10.1038/384285a0

48. Duman JG, Wu DW, Xu L, Tursman D, Olsen TM. Adaptations of insects to subzero temperatures. Q Rev Biol. 1991;66:387–410. https://doi.org/10.1086/417337

49. Sun X, Griffith M, Pasternak JJ, Glick B. Low temperature growth, freezing survival and production of antifreeze protein by the plant growth promoting rhizobacterium *Pseudomonas putida* GR12–2. Can J Microbiol. 1995;41:776–784. https://doi.org/10.1139/m95-107

50. Duman JG, Olsen TM. Thermal hysteresis protein activity in bacteria, fungi and phylogenetically diverse plants. Cryobiology. 1993;30:322–328. https://doi.org/10.1006/cryo.1993.1031

51. Yamashita Y, Nakamura N, Omiya K, Nishikawa J, Kawahara H, Obata H. Identification of an antifreeze lipoprotein from *Moraxella* sp. of Antarctic origin. Biosci Biotechnol Biochem. 2002;66:239–247. https://doi.org/10.1271/bbb.66.239

52. Muryoi N, Sato M, Kaneko S, Kawahara H, Obata H, Yaish MW, et al. Cloning and expression of *afp4*, a gene encoding an antifreeze protein from the arctic plant growth-promoting rhizobacterium *Pseudomonas putida* GR12-2. J Bacteriol. 2004;186:5661–5671. https://doi.org/10.1128/JB.186.17.5661-5671.2004

53. Krembs C, Eicken H, Junge K, Deming JW. High concentrations of exopolymeric substances in Arctic winter sea ice: Implications for the polar ocean carbon cycle and cryoprotection of diatoms. Deep Sea Res. 2002;49:2163–2181. https://doi.org/10.1016/S0967-0637(02)00122-X

54. Mancuso Nichols CA, Guezennec J, Bowman JP. Bacterial exopolysaccharides from extreme marine environments with special consideration of the southern ocean, sea ice, and deep-sea hydrothermal vents: A review. Mar Biotechnol. 2005;7:253–271. https://doi.org/10.1007/s10126-004-5118-2

55. Ko R, Smith LT, Smith GM. Glycine betaine confers enhanced osmotolerance and cryotolerance on *Listeria monocytogenes*. J Bacteriol. 1994;176:426–431. https://doi.org/10.1128/jb.176.2.426-431.1994

56. Phadtare S. Recent developments in bacterial cold-shock response. Curr Issues Mol Biol. 2004;6:125–136.

57. Tibbett M, Sanders FE, Cairney JWG. The effect of temperature and inorganic phosphorus supply on growth and acid phosphatase production in arctic and temperate strains of ectomycorrhizal *Hebeloma* spp. in axenic culture. Mycol Res. 1998;102:129–135. https://doi.org/10.1017/S0953756297004681

58. Hebraud M, Potier P. Cold shock response and low temperature adaptation in psychrotrophic bacteria. J Mol Microbiol Biotechnol. 1999;1:211–219.

59. Chaikam V, Karlson DT. Comparison of structure, function and regulation of plant cold shock domain proteins to bacterial and animal cold shock domain proteins. BMB Rep. 2010;43:1–8. https://doi.org/10.5483/BMBRep.2010.43.1.001

60. Phadtare S. Unwinding activity of cold shock proteins and RNA metabolism. RNA Biol. 2011;8:394–397. https://doi.org/10.4161/rna.8.3.14823

61. Kawahara H, Koda N, Oshio M, Obata H. A cold acclimation protein with refolding activity on frozen denatured enzymes. Biosci Biotechnol Biochem. 2000;64:2668–2674. https://doi.org/10.1271/bbb.64.2668

62. Samie N, Noghabi K, Gharegozloo Z, Zahiri H, Ahmadian G, Sharafi H, et al. Psychrophilic α-amylase from *Aeromonas veronii* NS07 isolated from farm soils. Process Biochem. 2012;47:1381–1387. https://doi.org/10.1016/j.procbio.2012.05.007

63. Struvay C, Feller G. Optimization to low temperature activity in psychrophilic enzymes. Int J Mol Sci. 2012;13:11643–11665. https://doi.org/10.3390/ijms130911643

64. Gerday C, Aittaleb M, Bentahier M, Chessa JP, Claverie P, Collins T, et al. Cold-adapted enzymes: From fundamentals to biotechnology. Trends Biotechnol. 2000;18:103–107. https://doi.org/10.1016/S0167-7799(99)01413-4

65. Miyazaki K, Wintrode PL, Grayling RA, Rubingh DN, Arnold FH. Directed evolution study of temperature adaptation in a psychrophilic enzyme. J Mol Biol. 2000;297:1015–1026. https://doi.org/10.1006/jmbi.2000.3612

66. Siddiqui KS, Cavicchioli R. Cold-adapted enzymes. Annu Rev Biochem. 2006;75:403–433. https://doi.org/10.1146/annurev.biochem.75.103004.142723

67. Aghajari N, Feller G, Gerday C, Haser R. Structures of the psychrophilic *Alteromonas haloplanctis* alpha-amylase give insights into cold adaptation at a molecular level. Structure. 1998;6:1503–1516. https://doi.org/10.1016/S0969-2126(98)00149-X

68. Humphrey DR, George A, Black GW, Cummings SP. *Flavobacterium frigidarium* sp. nov., an aerobic, psychrophilic, xylanolytic and laminarolytic bacterium from Antarctica. Int J Syst Evol Microbiol. 2001;51:1235–1243. https://doi.org/10.1099/00207713-51-4-1235

69. Bendt A, Huller H, Kammel U, Helmke E, Schweder T. Cloning, expression and characterization of a chitinase gene from the Antarctic psychrotolerant bacterium *Vibrio* sp. strain Fi:7. Extremophiles. 2001;5:119–126. https://doi.org/10.1007/s007920100179

70. Lu M, Fang Y, Li H, Liu H, Wang S. Isolation of a novel cold-adapted amylase-producing bacterium and study of its enzyme production conditions. Ann Microbiol. 2010;60:557–563. https://doi.org/10.1007/s13213-010-0090-8

71. Krishnan A, Convey P, Gonzalez-Rocha G, Alias S. Production of extracellular hydrolase enzymes by fungi from King George Island. Polar Biol. 2016;39:65–76. https://doi.org/10.1007/s00300-014-1606-7

72. Arora P. *Staphylococcus lipolyticus* sp. nov., a new cold-adapted lipase producing marine species. Ann Microbiol. 2013;63:913–922. https://doi.org/10.1007/s13213-012-0544-2

73. Huston AL, Methe B, Deming JW. Purification, characterization, and sequencing of an extracellular cold active aminopeptidase produced by marine psychrophile *Colwellia psychrerythraea* strain 34H. Appl Environ Microbiol. 2004;70:3321–3328. https://doi.org/10.1128/AEM.70.6.3321-3328.2004

74. Tsuruta H, Tamura J, Yamagata H, Aizono Y. Specification of amino acid residues essential for the catalytic reaction of cold-active protein-tyrosine phosphatase of a psychrophile, *Shewanella* sp. Biosci Biotechnol Biochem. 2004;68:440–443. https://doi.org/10.1271/bbb.68.440

75. Cavicchioli R, Charlton T, Ertan H, Mohd Omar S, Siddiqui K, Williams T. Biotechnological uses of enzymes from psychrophiles. Microbial Biotechnol. 2011;4:449–460. https://doi.org/10.1111/j.1751-7915.2011.00258.x

76. Banerjee R, Halder A, Natta A. Psychrophilic microorganisms: Habitats and exploitation potentials. Eur J Biotechnol Biosci. 2016;4:16–24.

77. Margesin R, Schinner F. Biodegradation and bioremediation of hydrocarbons in extreme environments. Appl Microbiol Biotechnol. 2001;56:650–663. https://doi.org/10.1007/s002530100701

78. Zhou Z, Qing J, Jinlai M, Fangming L. Antarctic psychrophile bacteria screening for oil degradation and their degrading characteristics. Mar Sci Bull. 2008;10:50–57.

79. Whyte LG, Bourbonnière L, Bellerose C, Greer CW. Bioremediation assessment of hydrocarbon-contaminated soils from the high Arctic. Bioremediation J. 1999;3:69–79. https://doi.org/10.1080/10889869991219217

80. Ruberto L, Vazquez S, Lobalbo A, Mac Cormack W. Psychrotolerant hydrocarbon-degrading *Rhodococcus* strains isolated from polluted Antarctic soils. Antarct Sci. 2005;17:47–56. https://doi.org/10.1017/S0954102005002415

81. Aislabie J, Saul DJ, Foght JM. Bioremediation of hydrocarbon-contaminated polar soils. Extremophiles. 2006;10:171–179. https://doi.org/10.1007/s00792-005-0498-4

82. Lin X, Yang B, Shen J, Du N. Biodegradation of crude oil by an Arctic psychrotrophic bacterium *Pseudoalteromomas* sp. P29. Curr Microbiol. 2009;59:341–345. https://doi.org/10.1007/s00284-009-9440-9

83. Nam E, Ahn J. Antarctic marine bacterium *Pseudomonas* sp. KNOUC808 as a source of cold-adapted hydrolyzing enzyme. Braz J Microbiol. 2011;42:927–936. https://doi.org/10.1590/S1517-83822011000300011

84. Hoyoux A, Jennes I, Dubois P, Genicot S, Dubail F, Francois JM, et al. Cold-adapted β-galactosidase from the Antarctic psychrophile *Pseudoalteromonas haloplanktis*. Appl Environ Microbiol. 2001;67:1529–1535. https://doi.org/10.1128/AEM.67.4.1529-1535.2001

85. Adapa V, Ramya L, Pulicherla K, Rao K. Cold active pectinases: Advancing the food industry to the next generation. Appl Biochem Biotechnol. 2014;172:2324–2337. https://doi.org/10.1007/s12010-013-0685-1

86. Singh P, Hamid B, Ahmad Lone M, Ranjan K, Khan A, Chaurse V, et al. Evaluation of pectinase activity from the psychrophilic fungal strain *Truncatella angustata*-BPF5 for use in wine industry. Endocytobiosis Cell Res. 2012;22:57–61.

87. Tomova I, Stoilova-Disheva M, Lazarkevich I, Vasileva-Tonkova E. Antimicrobial activity and resistance to heavy metals and antibiotics of heterotrophic bacteria isolated from sediment and soil samples collected from two Antarctic islands. Front Life Sci. 2015;8:348–357. https://doi.org/10.1080/21553769.2015.1044130

88. Gesheva V, Vasileva-Tonkova E. Production of enzymes and antimicrobial compounds by halophilic Antarctic *Nocardioides* sp. grown on different carbon sources. World J Microbiol Biotechnol. 2012;28:2069–2076. https://doi.org/10.1007/s11274-012-1009-2

89. Bisht S. Cold active proteins in food and pharmaceutical industry [article on the Internet]. Biotech Articles. c2011 [cited 2017 Dec. 15]. Available from: http://www.biotecharticles.com/Biotechnology-products-Article/Cold-Active-Proteins-in-Food-and-Pharmaceutical-Industry-719.html

90. Jadhav V, Jamle M, Pawar P, Devare M, Bhadekar R. Fatty acid profiles of PUFA producing Antarctic bacteria: Correlation with RAPD analysis. Ann Microbiol. 2010;60:693-699. https://doi.org/10.1007/s13213-010-0114-4

91. Babu J, Ramteke PW, Thomas G. Cold active microbial lipases: Some hot issues and recent developments. Biotechnol Adv. 2008;26:457–470. https://doi.org/10.1016/j.biotechadv.2008.05.003

92. Maiangwa J, Ali M, Salleh A, Abd Rahman R, Shariff F, Leow T. Adaptational properties and applications of cold-active lipases from psychrophilic bacteria. Extremophiles. 2015;19:235-247. https://doi.org/10.1007/s00792-014-0710-5

93. Araujo R, Casal M, Cavaco-Paulo A. Application of enzymes for textile fibres processing. Biocatal. Biotransform. 2008;26:332–349. https://doi.org/10.1080/10242420802390457

Trends in behavioural ecology: Putting South African research in a global perspective

AUTHOR:
Aliza le Roux[1]

AFFILIATION:
[1]Department of Zoology and Entomology, University of the Free State, Qwaqwa Campus, Phuthaditjhaba, South Africa

CORRESPONDENCE TO:
Aliza le Roux

EMAIL:
lerouxa3@ufs.ac.za

DATES:
Received: 18 Sep. 2017
Revised: 14 Dec. 2017
Accepted: 17 Jan. 2018
Published: 30 May 2018

KEYWORDS:
animal behaviour; bibliometrics; Mammalia; research impact; ecology

HOW TO CITE:
Le Roux A. Trends in behavioural ecology: Putting South African research in a global perspective. S Afr J Sci. 2018;114(5/6), Art. #2017-0321, 6 pages. http://dx.doi.org/10.17159/sajs.2018/20170321

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

South Africa remains at the leading edge of scientific publishing on the African continent, yet few analyses of publication patterns exist outside the biomedical field. Considering the large number of protected areas and mammalian guilds within the country, I examined trends in South African ecological research as it pertains to the behaviour of mammals. I assessed the topics and taxonomic focus of mammalogists at South African institutes over the span of 15 years (2001–2015), and contrasted local research with the shifting focus of international behavioural research. This review of more than 1000 publications indicates that South African based researchers exhibit a strong tendency towards field-based research, as opposed to laboratory-centred experiments. In terms of topical focus, local ecologists place significant weight on the behavioural categories of mating, social and foraging behaviour – reflecting a global priority for these topics. This finding contrasts with an increased emphasis on animal cognition and communication research in the international research arena, including field-based studies on these themes. I make suggestions on how behavioural ecologists in South Africa can align themselves with global trends while also continuing to distinguish those facets that make South African behavioural ecology unique.

Significance:

- This review is the first of behavioural ecology in South Africa.
- Suggestions are made for where South African researchers can profitably shift research focus.
- International trends in behavioural ecology are highlighted.

Introduction

South Africa remains at the forefront of scientific publishing on the African continent, with Life Sciences contributing strongly to these research outputs.[1] Within this broader discipline, we also have the relatively unique advantage of direct access to intact mammalian guilds[2] inside well-managed protected areas[3] – a factor that has stimulated significant mammalogical research in the country. However, no broad meta-analysis of local publishing patterns currently exists, and by some accounts few South African researchers are investigating mammalian behaviour: within the Zoological Society of Southern Africa, the Ethology research group was short-lived[4], and in the past 15 years only one paper on mammalian behaviour has been published in South Africa's top multidisciplinary journal, the *South African Journal of Science*[5]. This is surprising, as evolutionary biology is thriving in South Africa[6] and behavioural ecology is by definition the integration of evolutionary biology with the observation of animal behaviour. Since 1963[7], when Tinbergen created his seminal framework for studying behaviour (traditional 'ethology') in an evolutionary context, the study of behavioural ecology has thrived internationally[8]. How well, then, is South African research represented within this field?

I reviewed the study of mammalian behaviour in South Africa, describing trends in behavioural ecology over the course of 15 years (2001–2015) and highlighting the evolution of this discipline. As a framework, I use the 10 themes of behavioural ecology recently developed by Berger-Tal and colleagues[9] as an accurate reflection on the specialities within this research field. I discuss specific strengths and weaknesses of behavioural ecology research in South Africa and indicate opportunities for continued growth.

Methodology

I conducted an extensive survey of international literature on mammalian behavioural ecology using the SCOPUS database. Search terms were deliberately broad to encompass all literature (articles/review papers) from 2001 to 2015 that included the term "behav*" and "mammal*" in the title, abstract or keywords. In addition, I constructed a list of all mammalian families, thereby creating a search phrase to check if any of the mammalian families were included in the title, abstract or keywords of the publication. An initial search specifying these broad terms yielded 11 330 articles and review papers. This figure is an indication of the number of publications that describe behaviour, but is likely an exaggeration of the research that deals specifically and exclusively with mammalian behaviour.

I refined my search to compare South African research with the international literature, and created two publication databases. In the first, I specified that the author affiliation had to include "South Africa". These authors may have had multiple affiliations, and affiliation with a South African institute did not necessarily imply this was their primary place of work. I compiled a second database of international research that explicitly excluded authors with South African affiliations. In the second database, I limited the search to journals that focus specifically on animal behaviour: *Behavioral Ecology* (impact factor (IF)=3.177, from http://www.citefactor.org/ accessed on 24 November 2016), *Animal Behaviour* (IF=3.137), *Applied Animal Behaviour Science* (IF=1.691), *Advances in the Study of Behaviour* (IF=2.692), and *Behavioral Ecology and Sociobiology* (IF=2.35). Furthermore, I included publications on mammalian behaviour that featured in the top three multidisciplinary journals, to ensure that I did not ignore research with the putative highest impact. These journals were: *Science* (2013 IF=31.477), *Nature* (2013 IF=42.351), and *Proceedings of the National Academy of Sciences of the United States of America*

(2013 IF=9.423). Once the articles were extracted from the SCOPUS database, I analysed all abstracts to exclude papers that did not focus on, or explicitly include mammalian behavioural ecology. This process produced a total of 371 South African papers and 901 international publications. This refined selection of South African papers suggests that at least 3.3% of all articles on mammalian behaviour include South African affiliations (371 out of 11 330). Importantly, as I was selective in the outlets for international publications in particular, the latter database did not constitute all international publications on mammalian behaviour – it is a reflection of the articles appearing in the more respected and subject-specific journals, and therefore indicative of broader trends in the field.

Within these databases, I examined the taxonomic breadth of study species as well as the focus on different techniques or approaches in the local and international literature. To aid the thematic analyses, I performed text mining to identify the most common keywords, using the 'tm' package[10] in RStudio for Mac (version 0.99.903). Finally, I pooled the papers for each database into three 5-year blocks: (1) 2001–2005; (2) 2006–2010; and (3) 2011–2015, and used the 10 behavioural themes identified in a recent review[9] to guide my assessment of changing trends.

Findings and discussion

Overview: South African contributions

South African researchers contributed to >3% of the global literature on mammalian behavioural ecology, which is higher than the 1% contribution that sub-Saharan African countries make to global scientific research in general.[11] As the original database of articles on mammalian behaviour was likely an exaggerated number (I did not go through all 11 330 abstracts to ascertain each study's focus), the true proportion of South African affiliations will be even higher. The majority of South African papers were primary research articles, with local authors publishing review papers at a somewhat higher rate (2.5% of the South African database) than international authors (1.5% of the international database). While it is positive to see this contribution to the synthesis of the field, South African ecologists appear to have had a smaller impact than international scientists: only 13.2% of our research has been cited more than five times per year, in contrast to nearly one third (31.7%) of international articles being cited at the same rate. The South African ecologists who publish in top multidisciplinary journals (e.g. Slotow et al.[12]) often base their research on long-term field data, that is, studies that continue beyond a three-year project.

While many local and international publications claimed to investigate behaviour in an evolutionary context, South African researchers were more likely to address the ultimate causes and consequences of behaviour (in contrast to proximate or mechanistic causes of behaviour). South African authors tend to study species within the broader ecological context, incorporating aspects of the habitat[13], ecosystem[14], and the survival/fitness value[15] of different behaviours. Population dynamics and population-level analyses were common in South African research, often including genetic analyses of population structure.[16] In one third (36.4%) of South African papers[17,18], authors linked behaviour to conservation, compared with a much more limited focus on conservation in the international literature (1.8% of the literature, e.g. Palphramand et al.[19]). South African researchers therefore appear to align themselves well with calls for the integration of behaviour and conservation.[9] Furthermore, many South African researchers conducted experiments in their behavioural research, with nearly half (46.1%) of experiment-driven articles based on field studies.[20,21]

By contrast, in the international literature, 83.7% of experiments took place in captive or laboratory-based settings[22], and a further 5.4% of these experiments utilised domestic or farmed animals[23]. International scientists often followed a mechanistic approach to the analysis of mammalian behaviour, with a large proportion of publications investigating causes and consequences of behaviour on a cellular level.[24] These authors also frequently studied the mechanisms of circadian rhythms[25] and echolocation[26], which are largely absent from the South African database. International researchers exhibited a high interest in animal welfare research[27], likely reflecting the prevalence of captive species in their taxonomic repertoire.

The most popular taxonomic clade in both databases was Rodentia; however, international studies concentrated on laboratory rodents[28], contrasting with a much broader focus in South African research (including mole rats[14] and *Rhabdomys*[29] species). Carnivores were also prominent study subjects, with meerkats (*Suricata suricatta*) dominating both databases.[30,31] Internationally, canids[32] and spotted hyenas (*Crocuta crocuta*[33]) featured strongly in behavioural research, whereas South African researchers more often investigated Felidae[34]. Antelope[35] and small carnivores outside the Herpestidae family[36] were amongst the least studied taxa.

Dominant themes: Foraging, reproduction and social behaviour

Research in most thematic areas increased over the review period (Table 1), signifying a growing trend for researchers to address multiple themes within one study. South African and international authors largely agreed in terms of the top three thematic areas: foraging, reproductive and social behaviour (Table 1, Figure 1). In contrast to international trends, in which indirect analysis of foraging behaviour was rarely represented[37], many South African researchers use indirect methods such as scat[38] and isotope analyses[39] to describe foraging behaviour. Foraging ecology in the South African literature is therefore often restricted to dietary studies, with nuanced assessments of foraging behaviour – such as habitat selection[13] or strategic responses to environmental variation[40] – remaining uncommon.

Locally and internationally, research on mating and reproductive behaviour often assesses the hormonal correlates of reproductive success and/or helping behaviour. While common in the international literature[41], the endocrinology of reproduction is particularly widespread in the South African database (for example see Marneweck et al.[15]). In both databases, much of the research focuses on reproductive control or suppression[42], including rarely documented abortion[43].

Within the theme of social behaviour, relatively few researchers examined social behaviour and social structure as a goal in itself. Sociality was nearly always linked to other aspects of an animal's behavioural ecology, such as anti-predator and foraging behaviour[44], or competition[45]. The social behaviour of solitary species or solitary foragers was also examined fairly often by both local[46] and international authors[47]. The progression of this theme, beyond the basic description of social systems in gregarious species, is likely a reflection of the maturity of sociobiology as a research discipline.

Discrepancies: Communication, cognition and behavioural syndromes

A sharp contrast emerged between South African and international literature in terms of the next two high-priority themes. Internationally, communication and cognition were topics that featured in 17.7% and 16.1% of all articles, respectively. Comparatively speaking, South African authors largely have ignored these themes in the past 15 years, with only 2.7% of papers focused on learning and cognition, and 6.5% of publications investigating mammalian communication. The few South African publications examining mammalian cognition follow international trends in terms of topic, highlighting innovative behaviour in wild mammals[48], and exploring the link between stress and cognition[49]. One unusual study concluded that large brain size in dolphins was driven less by cognitive demands, than by low water temperatures.[50] This South African study stands in sharp contrast to seminal international articles linking brain size to sociality[51] and – more rarely – the demands of complex foraging[52]. Within the theme of mammalian communication, South Africans agree with international research in preferentially studying acoustic signals[53], above other communication modalities[54].

Over the review period, the theme of animal personality, or behavioural syndromes, was not investigated even once by South African mammalogists. The bulk of international publications (n=14) on the topic of personality appeared after 2010, although no papers on the theme appeared in the multidisciplinary journals *Science* or *Nature*. Mammalian personality or temperament is most often examined in the context of stress response[55], survival[56] and fitness[57].

Table 1: The percentage of papers published within each behavioural theme, across three 5-year blocks. South African papers include at least one South African author, whereas international papers were those works published without South African (co-)authors.

Theme	%South African papers			%International papers		
	2001–2005	2006–2010	2011–2015	2001–2005	2006–2010	2011–2015
Foraging	26.2	35.5	32.1	23.04	23.69	26.62
Mating/reproduction	24.6	33.3	32.1	29.57	35.54	37.74
Social	20.0	30.5	25.5	25.65	27.55	33.44
Movement	6.2	12.8	13.3	9.57	6.06	11.69
Competition	9.2	7.1	12.7	11.74	17.91	19.48
Parental care	10.8	11.3	9.7	4.78	14.88	11.69
Communication	1.5	6.4	8.5	16.52	18.18	17.86
Learning/cognition	0	2.1	4.2	16.52	14.33	17.86
Anti-predator	0	4.3	3.6	8.26	6.89	7.14
Personality	0	0	0	0.43	0.55	3.90

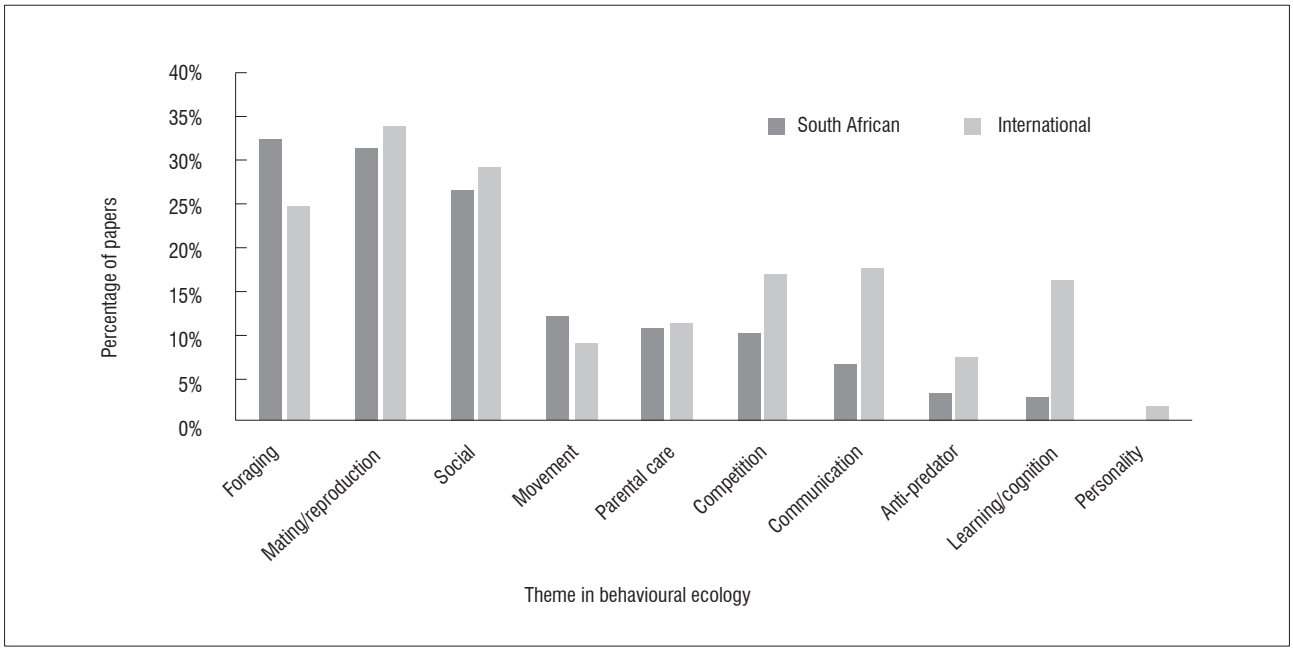


Figure 1: The percentage of articles and review papers examining 10 different themes within the discipline of behavioural ecology. South African authors' outputs between 2001 and 2015 are contrasted with the papers produced by non-South African (international) researchers over the same period.

The discrepancies between South African and international trends are less severe in the remaining five themes, and both databases present a similar distribution of topics. There are a few subtle distinctions in focus area. In terms of competitive behaviour, South African authors examine interspecific interactions[58] more often than international scientists do[59]. Within the theme of anti-predator behaviour, it is interesting to note that only international researchers appeared to consider the consequences of humans and human infrastructure for anti-predator behaviour[60], while vigilance[61] remains a common focus across both databases. I could find no remarkable distinctions between South African and international researchers in the study of mammalian migration and dispersal. Finally, in terms of parental care, the primary focus of both data sets was on maternal care, with South Africans[48] assessing paternal care more often than international authors[62].

The way forward

In a very practical sense, South African mammalogists are making optimal use of the accessible 'walk-in' laboratory available in the country's conserved areas and relatively intact ecosystems. A prominent focus on conservation-related research suggests that these researchers are cognisant of the risk of losing these fauna, and are actively directing projects to meet the challenges of a human-dominated global landscape. It is important to note, however, that using 'conservation' as a key term in research does not imply that practical solutions to conservation-

related problems are presented. Indeed, other reviewers[9] have pointed out that researchers who wish to have a positive impact on conservation challenges need to consciously conduct research that proposes, examines and practically explains answers to conservation-related questions. This is unfortunately not currently the norm in South African (or international) conservation biology.

Whereas behavioural ecology in South Africa is currently on firm footing, local researchers' impact can be more profound if we consciously adopt a more integrated approach. In a recent review, Bateson and Laland[8] highlighted that behavioural ecologists tend to ask specific behavioural questions in isolation, rarely combining a proximate and ultimate angle within a single study. This challenge remains despite the fact that Tinbergen[7] advocated for integration over 50 years ago, and South African researchers appear to be as guilty of this narrow theoretical scope as international researchers are. We could start addressing this challenge by building on our strengths. In the South African literature, the most cited research often addresses movement and foraging ecology (predation, in particular), and we have a strong national focus on studying carnivores. We could – and should – move beyond the study of movement and foraging behaviour (see also Young and Shivik[63]), building on the unique strengths of this mammalian order. For example, carnivores exhibit paternal care – a rarity amongst mammals[64] – more frequently than other mammals, yet we know almost nothing about the physiological drivers and consequences of this behaviour[65]. New studies should examine movement and foraging behaviour from both a proximate and ultimate angle, assessing the physiological covariates of dispersal and breeding behaviour, as well as the longer-term consequences of these individual decisions for fitness and population dynamics. Furthermore, while South African researchers already display a positive tendency to study heterospecific interactions, these studies are often performed on either a population level (e.g. Codron et al.[39]), or much smaller scale[58]. It is likely that collaborative research across study sites would enable us to assess inter-specific competition and collaboration on both a fine and large scale.

South African researchers appear slow to respond to some changing international trends that could open up new research and funding avenues. Specifically, local mammalogists pay scant attention to the themes of mammalian communication and cognition, contrasting with international trends. Importantly, international interest in the topic of animal cognition is far higher than this review would suggest, considering that I excluded subject-specific journals such as *Animal Cognition* (IF=1.122) from the review. Further, studies on mammalian communication and cognition are prominent amongst the most-cited papers in the international literature. Animal communication and cognition are, in many ways, closely linked. Not only is communicative complexity deemed indicative of cognitive complexity[66], but researchers often use animal signals in experiments designed to examine particular aspects of animal cognition[67]. South African authors could benefit from looking beyond the topics that have sustained local behavioural ecology for a long time, and steer some research focus into the minds and communicative abilities of our rich mammalian fauna. Of particular interest would be linking these themes with conservation biology.[68] For example, basic associative learning experiments can be used to train naïve animals to avoid predators[69] or avoid novel, toxic prey species[70]. Knowledge of how prey animals detect predators using olfactory, visual or acoustic cues could potentially improve the management of direct human–wildlife conflict: acoustic deterrents have been used to reduce crop raiding by elephants[71], and prey avoidance of predator faeces[72] suggests that such chemical cues can be used to deter some pests.

Another significant difference between South African and international research is that international scientists place a big emphasis on mechanistic studies of behaviour, often in captive study populations. This is particularly true for the behavioural research showcased in the top multidisciplinary journals, *Nature* and *Science*. Although this focus is likely a pragmatic response to the local paucity of wildlife, such controlled conditions open up novel research opportunities. The study of animal personality, for example, is largely rooted in laboratory or captive studies, and has only recently become more mainstream as a topic that has serious implications for animal survival and evolution.[73] There are endless possibilities for investigating the unique physiology and behaviour of our local fauna, particularly smaller species that may more readily adapt to captive environments. Already, some novel South African led research has emerged from research into captive mammals, such as an examination of how rising environmental temperatures impact sleep behaviour in bats (*Epomophorus wahlbergi*[74]). Certainly, local researchers could also tap into the international fascination with the mammalian brain, potentially through active collaboration with medical researchers who have access to high-end scanning equipment.

South African researchers' successful scientific exploitation of our abundant natural resources may have inadvertently created a research blindspot in that we literally do not look closer to home to investigate the myriad ways in which mammals respond to humans and human infrastructure. With humans often living inside or adjacent to protected areas in South Africa, and tourism being a primary source of national income, studying the direct and indirect interaction between humans or human infrastructure and wildlife is becoming imperative. At the moment, South African researchers have not looked much further than crop raiding or the negative impacts of primates 'invading' human-modified areas.[75,76] Thus, we typically concentrate on the negative impacts of human–wildlife on humans and largely ignore the impact of humans on the behaviour of local wildlife. Directions of research could include describing the behaviour of various species in urban areas[77], or assessing responses to humans as tourists[78], road users[60] and scientific observers[79].

South African behavioural ecologists would undeniably benefit from more collaboration between institutions, locations and fields of expertise. We still tend to work in silos (the majority of South African articles stemmed from the Mammal Research Institute of the University of Pretoria), and even concentrate on specific geographical locations, ignoring nature reserves that do not form part of the 'Big Five' national parks.[80] Funding bodies should also consider encouraging long-term field projects, which generate insights into the evolution of behaviour that no short-term approach could yield. Simultaneously, South African researchers should not ignore the possibilities inherent in well-designed laboratory and captive research that enable the detailed assessment of proximate corollaries of animal behaviour. We will not, however, make any great leaps forward without conversation. The time may be right for the revival of the Ethology research group as part of the Zoological Society of Southern Africa.

Acknowledgements

I am grateful to the University of the Free State's Office of the Research Directorate for research leave that enabled me to write this paper. I would also like to thank Diana Breshears for invaluable feedback on earlier drafts of this manuscript.

References

1. Pouris A. A bibliometric assessment of energy research in South Africa. S Afr J Sci. 2016;112:69–76. https://doi.org/10.17159/sajs.2016/20160054
2. Cameron EZ, Du Toit JT. Winning by a neck: Tall giraffes avoid competing with shorter browsers. Am Nat. 2007;169:130–135. https://doi.org/10.1086/509940
3. Licht DS, Kenner BC, Roddy DE. A comparison of the South African and United States models of natural areas management. ISRN Biodivers. 2014;2014:1–7. https://doi.org/10.1155/2014/737832
4. Taylor P, Hamer M. Standing on the shoulders of colourful giants: 50 years of zoological research in southern Africa. Afr Zool. 2009;44:217–231. https://doi.org/10.1080/15627020.2009.11407455
5. Midgley JJ, Balfour D, Kerley GI. Why do elephants damage savanna trees? S Afr J Sci. 2005;101:213–215.
6. Johnson SD. Darwin' s legacy in South African evolutionary biology. S Afr J Sci. 2009;105:403–409.
7. Tinbergen N. On aims and methods of ethology. Z Tierpsychol. 1963;20:410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

8. Bateson P, Laland KN. Tinbergen's four questions: An appreciation and an update. Trends Ecol Evol. 2013;28:712–718. https://doi.org/10.1016/j.tree.2013.09.013

9. Berger-Tal O, Blumstein DT, Carroll S, Fisher RN, Mesnick SL, Owen MA, et al. A systematic survey of the integration of animal behavior into conservation. Conserv Biol. 2016;30:744–753. https://doi.org/10.1111/cobi.12654

10. Feinerer I, Hornik K, Meyer D. Text mining infrastructure in R. J Stat Softw. 2008;25:1–54. https://doi.org/10.18637/jss.v025.i05

11. Tijssen RJW. Africa's contribution to the worldwide research literature: New analytical perspectives, trends, and performance indicators. Scientometrics. 2007;71:303–327. https://doi.org/10.1007/s11192-007-1658-3

12. Slotow R, Van Dyk G, Poole J, Page B, Klocke A. Older bull elephants control young males. Nature. 2000;408:425–426. https://doi.org/10.1038/35044191

13. Somers MJ, Nel JAJ. Habitat selection by the Cape clawless otter (*Aonyx capensis*) in rivers in the Western Cape Province, South Africa. Afr J Ecol. 2004;42:298–305. https://doi.org/10.1111/j.1365-2028.2004.00526.x

14. Hagenah N, Bennett NC. Mole rats act as ecosystem engineers within a biodiversity hotspot, the Cape Fynbos. J Zool. 2013;289:19–26. https://doi.org/10.1111/j.1469-7998.2012.00958.x

15. Marneweck D, Cameron EZ, Ganswindt A, Dalerum F. Behavioural and endocrine correlates to the aardwolf mating system. Mamm Biol. 2015;80:31–38. https://doi.org/10.1016/j.mambio.2014.08.001

16. Fontaine MC, Baird SJ, Piry S, Ray N, Tolley KA, Duke S, et al. Rise of oceanographic barriers in continuous populations of a cetacean: The genetic structure of harbour porpoises in Old World waters. BMC Biol. 2007;5:30. https://doi.org/10.1186/1741-7007-5-30

17. Hucke-Gaete R, Osman LP, Moreno CA, Findlay KP, Ljungblad DK. Discovery of a blue whale feeding and nursing ground in southern Chile. Proc R Soc London B Biol Sci. 2004;271(suppl 4):S170–173. https://doi.org/10.1098/rsbl.2003.0132

18. McMahon CR, Bester MN, Hindell MA, Brook BW, Bradshaw CJ. Shifting trends: Detecting environmentally mediated regulation in long-lived marine vertebrates using time-series data. Oecologia. 2009;59:69–82. https://doi.org/10.1007/s00442-008-1205-9

19. Palphramand KL, Newton-Cross G, White PCL. Spatial organization and behaviour of badgers (*Meles meles*) in a moderate-density population. Behav Ecol Sociobiol. 2006;61:401–413. https://doi.org/10.1007/s00265-006-0268-z

20. Dufour CMS, Meynard C, Watson J, Rioux C, Benhamou S, Perez J, et al. Space use variation in co-occurring sister species: Response to environmental variation or competition? PLoS ONE. 2015;10, e0117750, 15 pages. https://doi.org/10.1371/journal.pone.0117750

21. Jackson C, Bernard RTF. Effects of supplementary food on the winter inhibition of reproduction in male and female four-striped field mice (*Rhabdomys pumilio*). Reprod Fertil Dev. 2005;17:393. https://doi.org/10.1071/RD04134

22. Charlton BD, Ellis WAH, Brumm J, Nilsson K, Fitch WT. Female koalas prefer bellows in which lower formants indicate larger males. Anim Behav. 2012;84:1565–1571. https://doi.org/10.1016/j.anbehav.2012.09.034

23. Tallet C, Veissier I, Boivin X. A note on the consistency and specificity of lambs' responses to a stockperson and to their photograph in an arena test. Appl Anim Behav Sci. 2006;98:308–314. https://doi.org/10.1016/j.applanim.2005.09.005

24. Dosmann A, Brooks KC, Mateo JM. Evidence for a mechanism of phenotypic integration of behaviour and innate immunity in a wild rodent: Implications for animal personality and ecological immunology. Anim Behav. 2015;101:179–189. https://doi.org/10.1016/j.anbehav.2014.12.026

25. Maywood ES, Chesham JE, O'Brien JA, Hastings MH. A diversity of paracrine signals sustains molecular circadian cycling in suprachiasmatic nucleus circuits. Proc Natl Acad Sci USA. 2011;108:14306–14311. https://doi.org/10.1073/pnas.1101767108

26. Voigt-Heucke SL, Taborsky M, Dechmann DKN. A dual function of echolocation: Bats use echolocation calls to identify familiar and unfamiliar individuals. Anim Behav. 2010;80:59–67. https://doi.org/10.1016/j.anbehav.2010.03.025

27. Olsson IAS, Westlund K. More than numbers matter: The effect of social factors on behaviour and welfare of laboratory rodents and non-human primates. Appl Anim Behav Sci. 2007;103:229–254. https://doi.org/10.1016/j.applanim.2006.05.022

28. Roberts SA, Davidson AJ, Beynon RJ, Hurst JL. Female attraction to male scent and associative learning: The house mouse as a mammalian model. Anim Behav. 2014;97:313–321. https://doi.org/10.1016/j.anbehav.2014.08.010

29. Raynaud J, Müller K, Schradin C. Experimental increase of testosterone levels in free-ranging juvenile male African striped mice (*Rhabdomys pumilio*) induces physiological, morphological, and behavioral changes. Gen Comp Endocrinol. 2012;178:108–115. https://doi.org/10.1016/j.ygcen.2012.04.028

30. MacLeod KJ, Clutton-Brock TH. Low costs of allonursing in meerkats: Mitigation by behavioral change? Behav Ecol. 2015;26:697–705. https://doi.org/10.1093/beheco/aru205

31. Graw B, Manser MB. The function of mobbing in cooperative meerkats. Anim Behav. 2007;74:507–517. https://doi.org/10.1016/j.anbehav.2006.11.021

32. Mettler AE, Shivik JA. Dominance and neophobia in coyote (*Canis latrans*) breeding pairs. Appl Anim Behav Sci. 2007;102:85–94. https://doi.org/10.1016/j.applanim.2006.03.012

33. Curren LJ, Linden DW, Heinen VK, McGuire MC, Holekamp KE. The functions of male-male aggression in a female-dominated mammalian society. Anim Behav. 2015;100:208–216. https://doi.org/10.1016/j.anbehav.2014.11.024

34. Hayward M, O'Brien J, Kerley G. Carrying capacity of large African predators: Predictions and tests. Biol Conserv. 2007;139:219–229. https://doi.org/10.1016/j.biocon.2007.06.018

35. Wronski T, Apio A, Plath M. The communicatory significance of localised defecation sites in bushbuck (*Tragelaphus scriptus*). Behav Ecol Sociobiol. 2006;60:368–378. https://doi.org/10.1007/s00265-006-0174-4

36. Damasceno J, Genaro G. Dynamics of the access of captive domestic cats to a feed environmental enrichment item. Appl Anim Behav Sci. 2014;151:67–74. https://doi.org/10.1016/j.applanim.2013.11.004

37. Christianson D, Creel S. Risk effects in elk: Sex-specific responses in grazing and browsing due to predation risk from wolves. Behav Ecol. 2008;19:1258–1266. https://doi.org/10.1093/beheco/arn079

38. Ramesh T, Downs CT. Diet of serval (*Leptailurus serval*) on farmlands in the Drakensberg Midlands, South Africa. Mammalia. 2015;79:399–407. https://doi.org/10.1515/mammalia-2014-0053

39. Codron J, Duffy KJ, Avenant NL, Sponheimer M, Leichliter J, Paine O, et al. Stable isotope evidence for trophic niche partitioning in a South African savanna rodent community. Curr Zool 2015;61:397–411. https://doi.org/10.1093/czoolo/61.3.397

40. Woolley L-A, Millspaugh JJ, Woods RJ, Van Rensburg SJ, Page BR, Slotow R. Intraspecific strategic responses of African elephants to temporal variation in forage quality. J Wildl Manage. 2009;73:827–835. https://doi.org/10.2193/2008-412

41. Nguyen N, Gesquiere LR, Wango EO, Alberts SC, Altmann J. Late pregnancy glucocorticoid levels predict responsiveness in wild baboon mothers (*Papio cynocephalus*). Anim Behav. 2008;75:1747–1756. https://doi.org/10.1016/j.anbehav.2007.09.035

42. Faulkes CG, Bennett NC. Family values: Group dynamics and social control of reproduction in African mole-rats. Trends Ecol Evol. 2001;16:184–190. https://doi.org/10.1016/S0169-5347(01)02116-4

43. Roberts EK, Lu A, Bergman TJ, Beehner JC. A Bruce effect in wild geladas. Science. 2012;335:1222–1225. https://doi.org/10.1126/science.1213600

44. Shrader AM, Post JF, Hagenah N, Bateman PW. Is a reduction in the individual vigilance of mothers a key evolutionary driver of group formation in white rhinos? Afr Zool. 2013;48:109–114. https://doi.org/10.1080/15627020.2013.11407573

45. Hodge SJ, Flower TP, Clutton-Brock TH. Offspring competition and helper associations in cooperative meerkats. Anim Behav. 2007;74:957–964. https://doi.org/10.1016/j.anbehav.2006.10.029

46. Bray TC, Bloomer P, O'Riain MJ, Bennett NC. How attractive is the girl next door? An assessment of spatial mate acquisition and paternity in the solitary Cape dune mole-rat, *Bathyergus suillus*. PLoS ONE. 2012;7, e39866, 8 pages. https://doi.org/10.1371/journal.pone.0039866

47. Tchabovsky AV, Popov SV, Krasnov BR. Intra- and interspecific variation in vigilance and foraging of two gerbillid rodents, *Rhombomys opimus* and *Psammomys obesus*: The effect of social environment. Anim Behav. 2001;62:965–972. https://doi.org/10.1006/anbe.2001.1833

48. Le Roux A, Beishuizen R, Brekelmans W, Ganswindt A, Paris M, Dalerum F. Innovative parental care in a myrmecophageous mammal. Acta Ethol. 2014;17:63–66. https://doi.org/10.1007/s10211-013-0157-1

49. De Vos A, O'Riain JM, Meyer MA, Kotze PG, Kock AA. Behavior of Cape fur seals (*Arctocephalus pusillus pusillus*) in relation to temporal variation in predation risk by white sharks (*Carcharodon carcharias*) around a seal rookery in False Bay, South Africa. Mar Mammal Sci. 2015;31:1118–1131. https://doi.org/10.1111/mms.12208

50. Manger PR. An examination of cetacean brain structure with a novel hypothesis correlating thermogenesis to the evolution of a big brain. Biol Rev. 2006;81:293. https://doi.org/10.1017/S1464793106007019

51. Reader SM, Laland KN. Social intelligence, innovation, and enhanced brain size in primates. Proc Natl Acad Sci USA. 2002;99:4436–4441. https://doi.org/10.1073/pnas.062041299

52. McLean AN. Cognitive abilities -- the result of selective pressures on food acquisition? Appl Anim Behav Sci. 2001;71:241–258. https://doi.org/10.1016/S0168-1591(00)00181-7

53. Kriesell HJ, Elwen SH, Nastasi A, Gridley T. Identification and characteristics of signature whistles in wild bottlenose dolphins (*Tursiops truncatus*) from Namibia. PLoS ONE. 2014;9, e106317, 13 pages. https://doi.org/10.1371/journal.pone.0106317

54. Jordan NR, Cherry MI, Manser MB. Latrine distribution and patterns of use by wild meerkats: Implications for territory and mate defence. Anim Behav. 2007;73:613–622. https://doi.org/10.1016/j.anbehav.2006.06.010

55. Ferrari C, Pasquaretta C, Carere C, Cavallone E, Von Hardenberg A, Réale D. Testing for the presence of coping styles in a wild mammal. Anim Behav. 2013;85:1385–1396. https://doi.org/10.1016/j.anbehav.2013.03.030

56. Rödel HG, Zapka M, Talke S, Kornatz T, Bruchner B, Hedler C. Survival costs of fast exploration during juvenile life in a small mammal. Behav Ecol Sociobiol. 2015;69:205–217. https://doi.org/10.1007/s00265-014-1833-5

57. Rangassamy M, Dalmas M, Féron C, Gouat P, Rödel HG. Similarity of personalities speeds up reproduction in pairs of a monogamous rodent. Anim Behav. 2015;103:7–15. https://doi.org/10.1016/j.anbehav.2015.02.007

58. Do Linh San E, Somers MJ. Mongooses on the move: An apparent case of interspecific cooperative vigilance between carnivores. S Afr J Wildl Res. 2006;36:201–203.

59. Wauters L, Gurnell J, Martinoli A, Tosi G. Interspecific competition between native Eurasian red squirrels and alien grey squirrels: Does resource partitioning occur? Behav Ecol Sociobiol. 2002;52:332–341. https://doi.org/10.1007/s00265-002-0516-9

60. Shannon G, Angeloni LM, Wittemyer G, Fristrup KM, Crooks KR. Road traffic noise modifies behaviour of a keystone species. Anim Behav. 2014;94:135–141. https://doi.org/10.1016/j.anbehav.2014.06.004

61. Schmitt MH, Stears K, Wilmers CC, Shrader AM. Determining the relative importance of dilution and detection for zebra foraging in mixed-species herds. Anim Behav. 2014;96:151–158. https://doi.org/10.1016/j.anbehav.2014.08.012

62. Buchan JC, Alberts SC, Silk JB, Altmann J. True paternal care in a multi-male primate society. Nature. 2003;425:179–181. https://doi.org/10.1038/nature01866

63. Young JK, Shivik JA. What carnivore biologists can learn from bugs, birds, and beavers: A review of spatial theories. Can J Zool. 2006;84:1703–1711. https://doi.org/10.1139/z06-178

64. Kleiman DG, Malcolm JR. The evolution of male parental investment in mammals. In: Gubernick DJ, Klopfer PH, editors. Parental care in mammals. New York: Plenum Publishing Corporation; 1981. p. 347–386. https://doi.org/10.1007/978-1-4613-3150-6_9

65. De Bruin R, Ganswindt A, Le Roux A. From killer to carer: Steroid hormones and paternal behaviour. Afr Zool. 2016;51:173–182. https://doi.org/10.1080/15627020.2016.1258327

66. Kaplan G. Animal communication. Wiley Interdisciplinary Reviews: Cognitive Science. 2014;5:661–677. https://doi.org/10.1002/wcs.1321

67. Frommolt K-H, Goltsman ME, Macdonald DW. Barking foxes, *Alopex lagopus*: Field experiments in individual recognition in a territorial mammal. Anim Behav. 2003;65:509–518. https://doi.org/10.1006/anbe.2003.2064

68. Blumstein DT, Fernandéz-Juricic E. A primer of conservation behavior. Sunderland, MA: Sinauer Associates; 2010.

69. Griffin AS, Evans CS, Blumstein DT. Learning specificity in acquired predator recognition. Anim Behav. 2001;62:577–589. https://doi.org/10.1006/anbe.2001.1781

70. Price-Rees SJ, Webb JK, Shine R. Reducing the impact of a toxic invader by inducing taste aversion in an imperilled native reptile predator. Anim Conserv. 2013;16:386–394. https://doi.org/10.1111/acv.12004

71. Hill CM, Wallace GE. Crop protection and conflict mitigation: Reducing the costs of living alongside non-human primates. Biodivers Conserv. 2012;21:2569–2587. https://doi.org/10.1007/s10531-012-0318-y

72. Belton LE, Ball N, Waterman JM, Bateman PW. Do Cape ground squirrels (*Xerus inauris*) discriminate between olfactory cues in the faeces of predators versus non-predators? Afr Zool. 2007;42:135–138. https://doi.org/10.1080/15627020.2007.11407388

73. Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ. Integrating animal temperament within ecology and evolution. Biol Rev. 2007;82:291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x

74. Downs CT, Awuah A, Jordaan M, Magagula L, Mkhize T, Paine C, et al.Too hot to sleep? Sleep behaviour and surface body temperature of Wahlberg's epauletted fruit bat. PLoS ONE. 2015;10, e0119419, 14 pages. https://doi.org/10.1371/journal.pone.0119419

75. Jackson TP, Mosojane S, Ferreira SM, Van Aarde RJ. Solutions for elephant *Loxodonta africana* crop raiding in northern Botswana: Moving away from symptomatic approaches. Oryx. 2008;42:83–91. https://doi.org/10.1017/S0030605308001117

76. Henzi SP, Brown LR, Barrett L, Marais AJ. Troop size, habitat use, and diet of chacma baboons (*Papio hamadryas ursinus*) in commercial pine plantations: Implications for management. Int J Primatol. 2011;32:1020–1032. https://doi.org/10.1007/s10764-011-9519-6

77. Widdows CD, Downs CT. A genet drive-through: are large spotted genets using urban areas for 'fast food'? A dietary analysis. Urban Ecosyst. 2015;18:907–920. https://doi.org/10.1007/s11252-015-0438-8

78. Tadesse SA, Kotler BP. Impact of tourism on Nubian ibex (*Capra nubiana*) revealed through assessment of behavioral indicators. Behav Ecol. 2012;23:1257–1262. https://doi.org/10.1093/beheco/ars110

79. Nowak K, Le Roux A, Richards SA, Scheijen CP, Hill RA. Human observers impact habituated Samango monkeys' perceived landscape of fear. Behav Ecol. 2014;25:1199–1204. https://doi.org/10.1093/beheco/aru110

80. Van Wilgen BW, Boshoff N, Smit IPJ, Solano-Fernandez S, Van der Walt L. A bibliometric analysis to illustrate the role of an embedded research capability in South African National Parks. Scientometrics. 2016;107:185–212. https://doi.org/10.1007/s11192-016-1879-4

Privacy and user awareness on Facebook

AUTHORS:
Phillip Nyoni[1]
Mthulisi Velempini[2]

AFFILIATIONS:
[1]Department of Information Systems, North-West University, Mafikeng, South Africa
[2]Department of Computer Science, University of Limpopo, Polokwane, South Africa

CORRESPONDENCE TO:
Mthulisi Velempini

EMAIL:
mvelempini@gmail.com

DATES:
Received: 27 Apr. 2017
Revised: 29 Aug. 2017
Accepted: 07 Dec. 2017
Published: 30 May 2018

KEYWORDS:
social networks; personal data; online profiling; third-party applications; online advertising

HOW TO CITE:
Nyoni P, Velempini M. Privacy and user awareness on Facebook. S Afr J Sci. 2018;114(5/6), Art. #2017-0103, 5 pages. http://dx.doi.org/10.17159/sajs.2018/20170103

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
North-West University

Users' privacy on social media platforms continues to be important as users face numerous threats to their personal data. Social media sites such as Facebook store large amounts of users' personal data which make such sites prime targets for hackers. Research has shown that users have been subjected to privacy attacks in which hacked personal data are sold to online marketers. These incidents have prompted the need to protect users' privacy against data theft by third parties. We investigated the privacy risks that social media users on Facebook face when online. The privacy awareness of regular users of Facebook was evaluated through the observation of their online activities. Facebook was selected as a case study because it is the largest and most popular social media platform in South Africa. A sample group of Facebook users was selected for this study based on their activeness (or frequency of posting, uploaded or liking) on the site. Findings indicate that users' personal data can be obtained as they are publicly available on Facebook. The implication of this finding is that users lack adequate awareness on protection tools designed to protect their personal data, and as a result, they risk losing their data and privacy.

Significance:

- This study serves as an assessment tool for the privacy and security features of the social media site Facebook. This assessment tool can help users of social media sites to evaluate their own behaviour and usage patterns on Facebook. It can also assist social media site designers in considering the effectiveness of current measures, which are designed to ensure that the privacy and safety of users are protected.

Introduction

Social media have attracted robust debate around user privacy as these sites store users' personal data online.[1,2] User-generated content is at the core of Facebook as users share their opinions, personal pictures, location, age or gender.[2] When users share personal data, they do so without an understanding of the risks involved.[2] They assume that Facebook is a trusted computing platform but that is not always the case.[2] For example, hackers can create false accounts or clone user accounts to steal personal data.[3]

Third-party applications such as games on Facebook also present a threat to users' personal data.[2,3] These applications can also be used to access sensitive data as they always attempt to access users' Facebook profiles. A users' privacy can then be violated through the third-party application which can publish content using the identity of users which may violate privacy.[4] Third-party applications can profile and track online users' activities.[1]

Criminals can also track the movements of users whenever users post their geo-location data on Facebook, and could break into users' properties when they are away on holiday.[5] Facebook has attempted to offer tools for protecting users' privacy but the awareness of users of these tools is still lacking.[2] It is necessary to highlight possible risks associated with such self-disclosure tools.[6] It is envisioned that increased privacy awareness may encourage users to secure their data.[6]

We evaluated users' awareness of their privacy on Facebook. Our aim was to highlight social media privacy risks by using Facebook as a case study. Facebook was selected as it is popular and has been associated with a number of documented incidents of privacy violations. The site also encourages users to search for other users' profiles and add them as 'friends', which may violate their privacy.[3] This open sharing of data is at the heart of this study.

Social media: Facebook

Facebook is one of the largest social media sites with 1.28 billion users.[7] There are 50.3 million Facebook users in Africa and 5.5 million users in South Africa – making South Africa the second largest nation of Facebook users in Africa after Egypt (with 13 million users).[8] The site operates by getting users to connect to each other based on their background or shared interests.[2] It also allows them to join groups that have the same likes. Each user signs up for an online profile which contains personal data on the user such as their name and email address.[2] Part of being on Facebook involves users posting status updates which inform others about what they are doing. These updates then appear on their friends' newsfeeds as well as atop the user's feed.[2] These data are available to anyone and are considered to be in the public domain.[9] Because of the type of information posted, it is possible for an attacker to collect and target users based on the personal information they share.[9]

The creator of Facebook has in the past expressed that privacy is not as important as the value that the site offers.[10] Personalised services and targeted advertising on Facebook rely on users' personal information.[10] Tailoring services based on personal information allows companies to segment potential customers and advertise their products.[10]

Previous studies have focused on the usage patterns of university students on Facebook and did not examine the privacy issues faced by these students on Facebook.[9] In this study, we highlight the online privacy issues that users of Facebook encounter and we suggest how these issues may be mitigated.

Personal data and Facebook

Personal data are data that can be linked to an individual such as location or utility bills.[11] Facebook relies on these data as it needs content that is user generated.[11] Users are willing to disclose very personal aspects of their lives such as holiday trips and recent job promotions.[11] The implications of these data being available include online marketers profiling users or cyber criminals obtaining information on users.[11] Personal data have a high potential for misuse if obtained wrongfully.[11]

A report by the advocacy group Security and Privacy in Online Social Networks[12] in 2015 found that Facebook tracked users' browsing histories, including users who no longer had an account and those who had opted to not be tracked by the site. These direct violations of privacy may lead to users' data being less secure on Facebook[12] and are why it is important for users to be in control of who has access to their data[4].

Another scam that has been perpetrated on Facebook involves criminals targeting young teenage users.[13] These young users often share personal details of a trip out of town or a holiday on Facebook (geo-location data)[13]; scammers then call the user's parents pretending to be the police and to have arrested the user in the exact location which they shared on Facebook.[13] The scammers appear to be legitimate as they also provide other information that they have obtained from the user's profile such as age, hometown and school.[13] These scammers then demand money for bail to be sent to a false account. If the parents do not verify their claims, they end up paying and the scam is successful.[13] That such scams are perpetrated using Facebook demonstrates how personal information can be used against users by criminals.[3,5,11]

Data sharing reveals important information about how users interact, which helps third parties to profile users.[11] It is possible for the government to spy on individuals online by accessing their Facebook data.[13]

Data sharing model

An information privacy model developed by Conger[14] lays out the types of relationships that exist between users, website operators (such as Facebook) and third parties (online marketers). This model gives a visual illustration of how personal data can be passed from users to the service provider and then passed onto third parties without user consent.[14] The model is shown in Figure 1.

Figure 1 shows how privacy can be violated through the sale of their personal data.[1] A lack of awareness of what information is stored about users and how it is used has led to researchers questioning Facebook's approach towards privacy.[14]

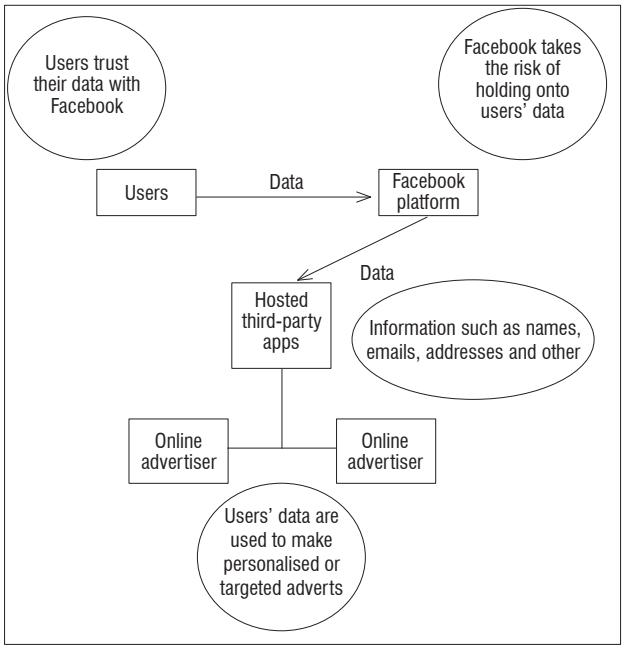


Figure 1: Information privacy model.[14]

Risks users face on Facebook

Risk is defined 'as a measure of uncertainty of an event happening times the severity of the outcome'[15]. Risk theory has been included here to explain why users may engage in unsafe behaviour online. Users may not be aware of potential threats to the data they post on Facebook. These potential threats include:

- Profiling. Big data analytics can be used against users by marketers or law enforcement agencies to profile them.[15]
- Scams and identity fraud. There have been a number of scams perpetrated on Facebook, from account cloning to users who impersonate officials for the purpose of defrauding individuals.[15]
- Surveillance and cyber bullying. The availability of personal data can be used against users for surveillance or harassment purposes.[15]

Disclosing personal information online has also affected some users in their search for employment.[11] Individuals are subject to background checks before signing employment contracts. These background checks involve reviewing social media accounts such as Facebook. Individuals who post and exhibit online behaviour that a prospective employer finds unprofessional could negatively affect their chances for employment.[11] Those already employed are at risk if they post any negative remarks about their organisation. These risks show how vulnerable personal information is and how users lack awareness of how to protect their personal data.[2] It is also possible that users engage in online self-disclosure as a consequence of ineffective privacy policies.[3]

Defining privacy

Privacy can be defined in a number of ways, but we adopted the definition provided by Westin[16]. Westin's[16] definition views privacy as the 'claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others'[16]. This definition is supported by Wacks[17] who describes privacy as the desire to be left alone. This view links privacy to the preservation of user identity for individuals.[17] It also highlights the need for users to control their own information, specifically how it is stored and disseminated by service providers.[18] This control can be implemented by giving users options for data minimisation such as a limited data sharing mode.[18] This option would allow users to preserve their privacy and grant them control over their data.[18]

Methodology

We utilised a mixed-methods approach for data collection. This approach was selected to add depth to the findings.[19] The methodology consisted of an online observation of users on Facebook (which constituted a natural setting). The users were observed using a polling checklist that gathered data from the profiles of users. In addition, a fake account was set up by the researcher to test how easy it is to clone a user's profile. Finally, a short survey was done on users' general awareness of privacy on Facebook. Ethical clearance for this study was given by the North-West University (NWU) Research Ethics Committee (reference number NWU-00212-13-A9).

Participants were drawn from NWU stakeholders. The study focused on Facebook because of its wide adoption in Africa and South Africa.[8] It was also considered to be ideal for the study given its publicly available and searchable profiles. The study targeted users who had liked the NWU Facebook page. These users included students, staff, alumni, prospective students, business associates and other stakeholders of the university. NWU was selected as a research site as it has a diverse number of individuals including African students and employees. The international students are approximately 6% of the student body. The findings of the research can be generalised to the broader community of African Facebook users.

Online observation procedure

The profile pages of users were compared against a polling checklist (see Appendix 1 in the supplementary material) that was organised according to different themes. In total, 357 profile pages were accessed based

on the convenience sample drawn from a population of 5701 users who liked the NWU Facebook page. This sample size was calculated using guidelines provided by Krejcie and Morgan[20]. Their guidelines help researchers find appropriately representative samples from target populations. Data collection took approximately 2 months in total and at least 15 minutes per user.

Facebook account cloning attack

To validate the results of the polling checklist, a fake Facebook profile page was created. The aim of this account cloning attack was to evaluate whether users were able to detect a false account trying to gain access to their account. The attack began by sending out friend requests from the fake account. Once the request was accepted, users were informed about the purpose of the attack. The personal information of the users who accepted the request was made available to the researcher for analysis. A total of 237 users were 'friended'.

User surveys

Two short user surveys were also conducted. The first survey was based on the polling checklist and the second on the account cloning attack. The surveys were done to support and validate the results of the previous methods. The first survey used convenience sampling to access participants from the population. Questionnaires were distributed to the research participants for completion and were collected as soon as the participants were done. A total of 25 individuals participated. The total number of responses was considered to be sufficient as this short survey was designed to validate the online observation results. The second survey was based on 30 third-year and honours students who volunteered to participate in a cyber security awareness training programme.

Results

The online observation phase of data collection was based on the users who had liked the NWU Facebook page. The sample population consists of 357 users of whom 55% (n=198) are women and 45% (n=159) are men. The most active users were within the 18–25 year age group (n=214); this finding was to be expected considering that the majority of students using Facebook are undergraduate students.

It was also found that 67% (n=240) of Facebook users' personal data are partially available, while 33% (n=117) have their full personal details available (Figure 2). Facebook does not put a default block on new users' personal information when they sign up to be a member on the site, which makes it easier for users to view each other's information, and also makes it possible for those with malicious intent to obtain sensitive data. Attackers seek out user names and passwords for Facebook by data mining those credentials. Other people use that information to deceive or market their products to the users through spam email.

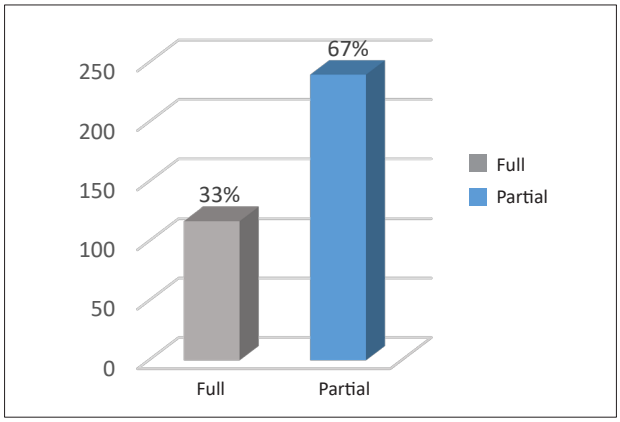


Figure 2: Availability of users' details.

Most users' data are partially available on Facebook, possibly because Facebook needs user profiles to be semi-accessible to the public in order for people to connect with users with common interests. The unfortunate side effect of this accessibility is that not everyone wants to be a friend on Facebook.

Most users (75%; n=269) either often share or sometimes share their geo-location with their friends on Facebook (Figure 3). Most of these users indicated that they share their location when they travel for holidays or when they spend time with friends. Users trust that their data are safe and share their daily activities on Facebook. Criminals can use this information to track users' movements and map their patterns, resulting in a high number of scams on Facebook.

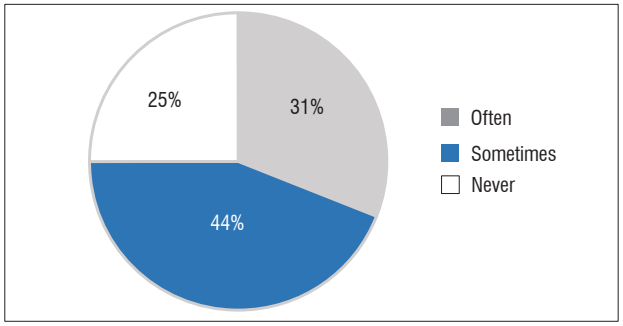


Figure 3: Geo-location sharing by users.

Figure 4 shows that 56% (n=202) of users post daily on Facebook, while 38% (n=135) post at least once or twice a week. Figure 4 also reveals that many users access Facebook through their mobile devices as smart devices have global-positioning sensors on them which can share location. Anyone can profile a user's daily routine from the frequency of their updates and location of their postings. Personal data are generated on a daily basis which makes it possible to track and profile such users.

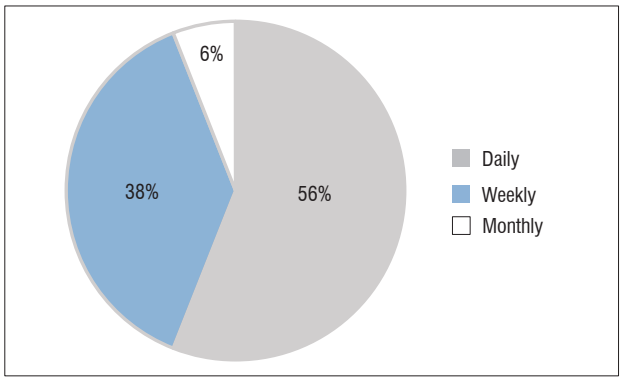


Figure 4: Frequency of user sharing.

The most common activity on Facebook is posting status updates (47%), as shown in Table 1; 14% uploaded pictures the most. In some cases, picture or video uploads were personal in nature and displayed a user's car registration number or house number. This practice is not exclusive to Facebook as other sites such as Instagram also have such images.

Table 1: Frequency of user activity

Activity	Frequency
Posting	169 users (47%)
Commenting	70 users (20%)
Liking	68 users (19%)
Uploading	50 users (14%)

Some users also post pictures of friends and Facebook's facial recognition feature tags them automatically without their consent.[13] These pictures can be digitally altered or used for cyber bullying (using the user's image for online jokes or memes) or for propaganda in the case of a public figure.[13] These practices can damage a user's reputation unless the user quickly un-tags themself from the image.

Facebook account cloning attack

An account was cloned and used to see if users could be lured by a fake account. Friend requests were sent out and as users responded, they were informed about the objective of this profile. The response rate to this page is shown in Figure 5. A total of 87 out of 237 users had accepted the invitation at the time the results were retrieved. This attack was run over the course of 1 month.

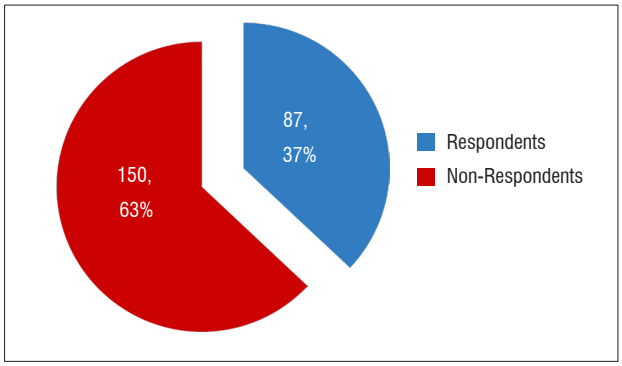


Figure 5: Response rate to a friend request from a fake account.

The users who responded did not verify the personal details to assess the veracity of the profile page. For example, users did not realise that the profile name and the name of the owner had been modified. It is common practice for Facebook users to either misspell their names purposely or use pseudo-names because they want to hide their identity, but this practice can also lead to users being tricked into accepting account impersonators. The attack indicates that a number of users on Facebook still lack privacy awareness.

User surveys

A short user survey was conducted to examine the privacy awareness of Facebook users. The respondents (n=25) confirmed that they had a Facebook profile and were active on it. Of the 25 respondents, 20 agreed that they shared personal data on Facebook. These data consisted of addresses and travel plans which could be exploited by attackers. Most respondents admitted that they frequently uploaded pictures, 13 changed their status regularly, 10 commented and 7 respondents shared their location often – a finding which supports the results of the online observation regarding geo-location sharing. The results of the survey are shown in Figure 6.

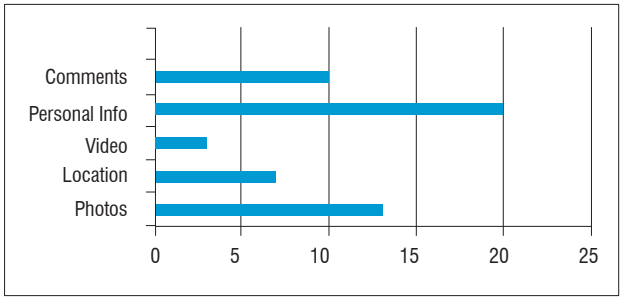


Figure 6: Information shared on users' profiles.

Figure 7 shows that 22 (88%) respondents never use Facebook privacy settings to protect their data. This may be because users do not know that these settings exist or may not know how to activate them, which may leave their personal data vulnerable to any potential profiling.

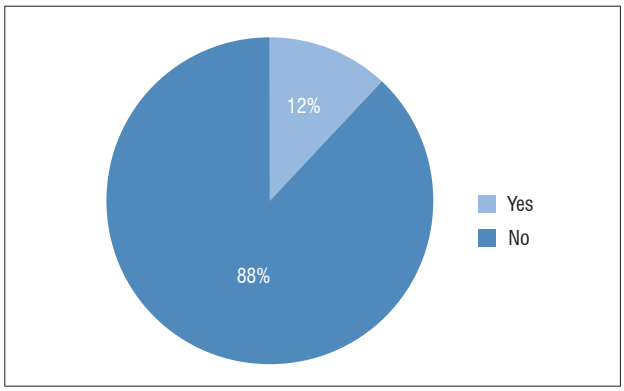


Figure 7: Privacy settings usage.

The second short survey was to investigate whether users were willing to meet someone they connected with on Facebook. A total population of 30 students were asked how they would respond to a request to meet in the real world. The results showed that 41% (combined from 33% and 7%) were willing to meet in person a Facebook friend who they had never met before (Figure 8). This willingness to trust a total stranger may lead to the users being defrauded or scammed by impersonators on Facebook.[5]

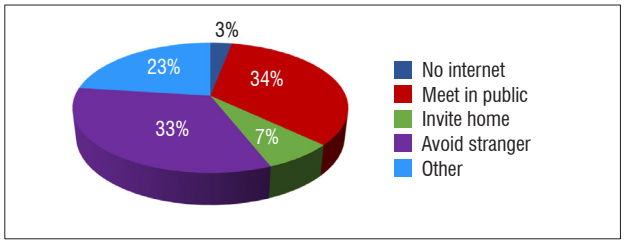


Figure 8: Willingness to meet with a stranger friended on Facebook.

Discussion

Based on the findings of this study, it is necessary for users to be trained on privacy settings on Facebook. Metadata (such as location) accompany posts and uploads that users create online and these 'extra data' can be used for surveillance or profiling purposes. While Facebook uses these metadata to tailor adverts that users see, they may also be misused by third parties. Someone could break into a user's home after obtaining the information on Facebook and studying their movement patterns from geo-location tags.

Facebook does have a comprehensive privacy policy in place to deal with some of these challenges. It covers issues such as how data are used, shared, viewed, changed, or removed.[21] Facebook also tries to elicit feedback from users concerning the policy in order to improve it and make it more effective.[21] However, the privacy policy is long and written in technical language which is not easily understood by most users. The policy highlights that privacy is a shared responsibility and users need to be proactive as well. Despite this policy, many users are not aware of this contractual obligation and do not use privacy settings to secure their data.

A conceptual model that reflects privacy and personal information on Facebook has been developed and is shown in Figure 9. It was developed using the findings of the online observation, account cloning attack and user surveys. The aim of this model is to highlight the roles and responsibilities of users, site providers (Facebook in this case) and third parties (i.e. online marketers).

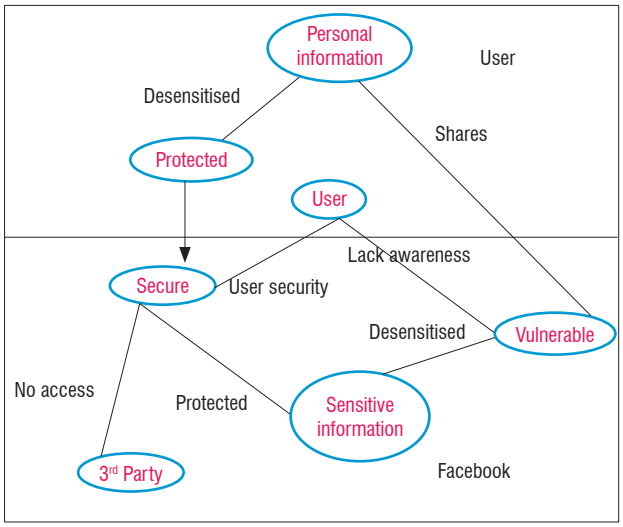


Figure 9: Model of the responsibilities of online actors.

The actors have a shared responsibility to protect and maintain the privacy of data. Users should make use of privacy settings to secure their data whenever they are online. Meanwhile Facebook is responsible for the provision of a secure platform and the enforcement of its privacy policy. Third parties must also ensure that personal data are not stolen or misused. Pro-activeness is necessary for each of these responsibilities to be achieved.

Conclusion

This study has revealed that users regularly post sensitive data, which can be used to track their movements and activities. Most users are not aware that their posts and updates are in the public domain and can be easily accessed. It is necessary to raise users' privacy awareness to protect them from possible loss of property or surveillance. Privacy settings on Facebook should be simplified for users to understand and given more emphasis so they are used. It is also important for laws that protect users' data to be enforced by regulators. Based on our findings, privacy awareness could be achieved through better user training on how to use privacy settings on Facebook. Users must be taught the different ways in which they can secure their personal information.

Acknowledgement

We thank North-West University for supporting this study.

Authors' contributions

P.N. was responsible for conceptualisation of the study, methodology, data collection, data analysis, sample analysis, validation, data curation, and writing the initial draft. M.V. was responsible for conceptualisation, methodology, student supervision, project leadership, critically reviewing the initial draft and the revisions, and acquiring the funding.

References

1. Titiriga R. Social transparency through recommendation engines and its challenges: Looking beyond privacy. Econ Inform J. 2010;15(4):147–155.
2. Kumar DV, Varma P, Pabboju SS. Security issues in social networking. Int J Comput Sci Netw Security. 2013;13(6):120–124.
3. Malik H, Malik AS. Towards identifying the challenges associated with emerging large scale social networks. Proc Comput Sci. 2011;5:458–465. https://doi.org/10.1016/j.procs.2011.07.059
4. Spinelli CF. Social media: No 'friend' of personal privacy. The Elon Journal of Undergraduate Research in Communications. 2010;1(2):59–69.
5. Blair K. New survey: Burglars use social media to plan crimes [webpage on the Internet]. c2011 [cited 2016 Nov 25]. Available from: http://socialtimes.com/new-survery-burglars-use-social-media-to-plan- crimes_b79475
6. Balduzzi M, Platzer C, Holz T, Kirda E, Balzarotti D, Kruegel C. Abusing social networks for automated user profiling. In: Jha S, Sommer R, Kreibich C, editors. Recent advances in intrusion detection. RAID 2010. Lecture Notes in Computer Science. 2010;6307:422–441. https://doi.org/10.1007/978-3-642-15512-3_22
7. Digital Insights. Social media statistics for 2014 [webpage on the Internet]. c2014 [cited 2016 Apr 14]. Available from: http://www.adweek.com/socialtimes/files/2014/06/social-media-statistics-2014.htm
8. Social Bakers. Africa Facebook users infographic [webpage on the Internet]. c2013 [cited 2016 Nov 20]. Available from: http://www.socialbakers.com/africa-facebook-users-infographic.jpg
9. Pempek TA, Yermolayeva YA, Calvert SL. College students' social networking experiences on Facebook. J Appl Dev Psychol. 2009;30(3):227–238. https://doi.org/10.1016/j.appdev.2008.12.010
10. Johnson B. Privacy no longer a social norm says Facebook founder [webpage on the Internet]. c2010 [cited 2016 Oct 01]. The Guardian. 2010 January 11. Available from: https://www.theguardian.com/technology/2010/jan/11/facebook-privacy
11. Furnell SM. Online identity: Giving it all away? Information Security Technical Report. 2010;15(2):42–46. https://doi.org/10.1016/j.istr.2010.09.002
12. Security and Privacy in Online Social Networks. From social media service to advertising network: A critical analysis of Facebook's revised policies and terms [document on the Internet]. c2015 [cited 2016 Sep 14]. Available from: https://www.law.kuleuven.be/icri/en/news/item/facebooks-revised-policies-and-terms-v1-2.pdf
13. Payton T, Claypoole T. Privacy in the age of big data. Lanham: Rowman & Littlefield; 2014.
14. Conger S. Emerging technologies, emerging privacy issues. In: Luppicini R, Adell R, editors. Handbook of research on technoethics. Hershey, PA: IGI Global; 2009. p. 767–793. https://doi.org/10.4018/978-1-60566-022-6.ch050
15. Riesch H. Levels of uncertainty. In: Handbook of risk theory. Amsterdam: Springer; 2012. p. 87–110. https://doi.org/10.1007/978-94-007-1433-5_4
16. Westin A. Privacy and freedom. New York: IG Publishing; 1967. p. 15–20.
17. Wacks R. Privacy: A very short introduction. New York: Oxford Press; 2010. https://doi.org/10.1093/actrade/9780199556533.003.0001
18. Ellison N, Vitak J, Steinfield C, Gray R, Lampe C. Negotiating privacy concerns and social capital needs in a social media environment. In: Trepte S, Reinecke L, editors. Privacy online. Berlin: Springer; 2010. p. 19–32.
19. Hesse-Biber SN. Mixed methods research: Merging theory with practice. New York: Guilford; 2010.
20. Krejcie R.V, Morgan D.W. Determining sample size for research activities. J Educ Psychol Measure. 1970;30(608):56. https://doi.org/10.1177/001316447003000308
21. Facebook. Privacy policy of Facebook [webpage on the Internet]. No date [updated 2014; cited 2016 Apr 14]. Available from: http://www.facebook.com/policies/privacy/basic/?ref_component

Cultural differences and confidence in institutions: Comparing Africa and the USA

AUTHOR:
Bankole Falade[1]

AFFILIATION:
[1]Department of Psychological and Behavioural Sciences, London School of Economics and Political Science, London, United Kingdom

Current address: South African Research Chair in Science Communication, Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Bankole Falade

EMAIL:
bankolefalade@gmail.com

DATES:
Received: 23 Apr. 2017
Revised: 15 Nov. 2017
Accepted: 07 Dec. 2017
Published: 30 May 2018

KEYWORDS:
cross-cultural studies; cultural metrics; Nigeria; Zimbabwe; Ghana; factor analysis

HOW TO CITE:
Falade B. Cultural differences and confidence in institutions: Comparing Africa and the USA. S Afr J Sci. 2018;114(5/6), Art. #2017-0135, 8 pages. http://dx.doi.org/10.17159/sajs.2018/20170135

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
None

A comparison was undertaken of confidence in 17 institutions in Ghana, Nigeria, Zimbabwe and the USA using data from the World Values Survey to find shared valuations and distinguishing characteristics as markers of cultural categories. Frequencies and rankings were examined and exploratory factor analysis was used to find plausible meanings of groups of institutions. The findings show that, although African respondents score institutions higher than their US counterparts, the rankings vary. With frequencies, the meaning is manifest. The analysis shows that 10 institutions load similarly on one latent variable and their combinations with the others indicate culture-specific characteristics. The latent variables were named 'not-for-profit', 'for-profit', 'political', 'watchdog or fourth estate' and 'social order' and they show Ghana is closer to the USA than to Nigeria, which is closer to Zimbabwe. The 'not-for-profit' variable is more important in the USA and Ghana and 'political' is more important in Nigeria and Zimbabwe. Institutional-specific loadings show that whereas the police and courts are grouped as 'political' in Nigeria, in other countries they belong to 'social order'; and while universities are perceived as 'for-profit' in Africa, they are 'not-for-profit' in the USA. Comparing frequencies and rankings or dividing along the lines of individualistic versus collective or private and public sectors, masks the dynamic distribution of the systems of meaning in the local cultures; the latent variables approach therefore offers a more conceptually sound categorisation informed by shared and distinguishing institutions.

Significance:

- Nigerians, as at the time of the survey, were yet to perceive the principles of separation of powers between political institutions, the judiciary and the police – an essential feature of a good democracy and a characteristic of other countries in the study. Zimbabweans and Nigerians perceive their public institutions in generally the same way with the domination of the political establishments while Ghanaians are closer to the USA in terms of the values they attach to their establishments with the most important group being the charities. The universities in Africa, as well as the civil service in Nigeria, are associated with business/profit centres with the Nigerian labour movement also seen as political. The army also remains relevant as a part of the fourth estate in Ghana and Nigeria.

Background

African countries have in one generation been under colonial, democratic, minority and military regimes and some have fought civil wars.[1,2] The political, social, economic and cultural experiences of individual African countries has thus varied. While many were under colonial rule, others were governed by minority governments. During colonisation, all institutions were controlled by foreign governments. Under military dictatorships, separation of powers, freedom of speech and political association were limited or suspended by decrees. Under minority rule, rights of the black majority were severely curtailed. The experience of the African public with its institutions has thus varied across the countries and this study is about finding the underlying common sources of influence and distinguishing characteristics in comparison with those of a long-established democracy – the USA.

In many African countries, military establishment has been the unofficial opposition party, and at times alternated with the civilian government[3] or provided internal stability[4]; the July 2013 coup in Egypt[5] shows that military establishments are still politically active in Africa. Independence of the judiciary and police from regime influence is in doubt under military rule, as human rights and pro-democracy activists and journalists are detained, jailed or killed[6-8], and under civilian rule, as the outcome of national elections are contested[9,10].

Southern African countries were free of military rule but were under white minority governments until 1980 in Zimbabwe and 1994 in South Africa. During minority rule, freedom of speech and political association were curtailed; several public institutions and places were segregated and access to political power, education and wealth was limited for blacks.[11,12] There was, however, a 15-year guerrilla warfare with nationalist forces that engaged the white minority government in Zimbabwe[13] and the anti-apartheid resistance movement in South Africa had a military wing, the Umkhonto we Sizwe[14].

Case studies: Ghana, Nigeria, Zimbabwe and the USA

The African countries Ghana, Nigeria and Zimbabwe were compared against the USA. Between independence from Britain in 1960 and 1999, when Nigeria's last military rule ended, soldiers held administrative and legislative powers for 29 out of 39 years and in the period of democratic governance, the ruling party was not defeated at federal elections until 2015. From 1966 to 1992, there was a series of military governments in Ghana but since its return to democratic government in 1992, power has changed hands at the central government consistently between one political party and another. Zimbabwe has been free of military intervention in politics since independence in 1980 but it has been governed by only one political party and president. By contrast, the USA has had a much longer history of democratic governance and transfer of power among political parties with the military confined to its constitutional role.

The Hofstede[15] score (https://www.geert-hofstede.com/countries.html) of individualism shows that Ghana with a score of 15 is more collectivist than Nigeria with a score of 30. The more individualistic USA has a score of 91 and no record exists for Zimbabwe. In addition to being collectivist, religion, which is inherently cultural in nature and shapes peoples' psychologies in complex ways,[16,17] is also important in African cultures compared with their Western counterparts.

All three African countries are struggling with weak economies and – like other countries on the continent – corruption, which undermines confidence in institutions.[18] In 2011, of 183 countries measured on the corruption perception index of 0 (highly corrupt) to 10 (very clean), Ghana ranked 69th with an index of 3.9; Nigeria ranked 143rd with 2.4; Zimbabwe 152nd with 2.2; and the USA ranked 24th with an index of 7.1.[19] On the economic front, per capita income – an indicator of personal wealth – was comparatively low for the African countries. Per capita income in 2011 for Nigeria was USD2514, Ghana USD1590, Zimbabwe USD769 and the USA USD49 800.[20] Inflation figures, indicators of macro-economic stability,[21,22] in 2011 (the year under review) were 10.8% for Nigeria, 8.7% for Ghana, 3.2% for Zimbabwe and 3.1% for the USA. Zimbabwe, at the peak of an economic crisis in 2007, had an inflation rate of 24 411%.[23]

The experience of African publics and public perception of the roles of their institutions in society are thus very likely to be different across the continent and from their counterparts in the West. How these differences may act as markers of culture was explored in this study by comparing Ghana, Nigeria and Zimbabwe with the USA. For this research, the African countries are categorised as within culture while comparing with the USA is across cultures. The 17 institutions considered are churches, armed forces, the press, television, labour unions, police, courts, central government, political parties, parliament, civil service, universities, major companies, banks, environmental organisations, women's organisations and charitable or humanitarian organisations. The study uses data from Wave 6 of the World Values Survey (www.worldvaluessurvey.org).

Culture, trust and confidence in institutions

Culture and institutions

Confidence in institutions is an indicator of the underlying feelings of the public about its polity[24] and is partly determined by subjective well-being, political attitudes, values[25], the state of the economy, employment, etc.[26] Subjective well-being, ideas, attitudes, values and beliefs are non-material components of culture and the levels of confidence in institutions and patterns of interrelations among them, it is argued here, establishes categorisations which distinguish one culture from the other.

Culture consists of shared elements that provide the standards for perceiving, believing, evaluating, communicating and acting among those who share a language, a historical period and a geographical location.[27,28] It is an embodiment of shared values and categorisations.[29] Noting the distinction and possible conflict between 'inner' and 'outer' models of culture, Mascolo[30] proposes that culture be defined as a dynamic distribution of systems of meanings, practices and artefacts – an approach which underscores the multiplicity and location of meanings in culture.

Also, collectivism and individualism are polythetic constructs whose attributes define cultures.[31] Collectivism is maximal when a society is low in complexity and tight and individualism is greatest in societies that are loose and complex.[27] Collectivism is high in African societies noted for strong family and in-group bonds, strong adherence to norms and multilingualism compared with their Western counterparts which are more individualistic and often monolingual. Gorodnichenko and Roland[32] argue that countries with collectivist cultures are more likely to experience autocratic breakdowns and transitions from autocracy to autocracy.

Cultural mapping is important to identify shared elements and distinguish characteristics with a view to understanding how meaning systems are distributed[27,28,30] and to know the mechanisms that are implied when certain values are endorsed and others are 'frowned upon'[33]. The evolving patterns will enhance our understanding of the underlying feelings of the public[24] about these institutions, in the countries of interest and others.

Familiarity, confidence and trust

Familiarity, confidence and trust are different modes of asserting expectations but they use self-reference in different ways. Familiarity draws the distinction between familiar and unfamiliar fields and aligns with the familiar while confidence emerges in situations characterised by contingency and danger, which makes it meaningful to reflect on pre-adaptive and protective measures. Trust, however, depends not on inherent danger but on risk which emerges only as a component of decision and action. Trust and confidence refer to expectations which may lapse into disappointments; a relation of confidence may turn to one of trust and, conversely, trust can revert to mere confidence, but it is not a simple zero-sum game.[34]

Luhmann[34] argues that a distinction between confidence and trust is not obsolete, although they belong to the same family of self-assurances and seem to depend on each other and are, at the same time, capable of replacing each other to a certain extent. For Luhmann, in the case of confidence, attribution for disappointment is external while for trust, it is internal. Also, lack of confidence will lead to alienation while lack of trust reduces the range of possibilities for rational action but the withdrawal of trust is not an immediate and necessary result of a lack of confidence. Leaning on Luhmann, Siegrist and colleagues[35] propose that confidence is based on high levels of familiarity and can be had in just about anything while trust is important when familiarity is low and is directed at persons.

Trust and confidence in institutions

Authors use the terms faith, confidence and trust in institutions interchangeably.[26,36,37] Some pollsters and researchers use the terms to measure public feelings scaled as trust, confidence or faith in leaders, government and other institutions. Other researchers use them for different variables and at different levels. Twenge et al.[37] use trust at the individual level and confidence at the institutional level. Cook and Gronke[38] separated trust as that in government and confidence as that in institutions while Tyler[36] used public trust and confidence together in research on legal authorities. Hager and Hedberg[39] argue that institutional trust and sector confidence are different from each other but nested within a generalised social trust in unknown others. Again, is the distinction important, in particular, for research?

Lipset and Schneider[26] suggest that while the terms may be varied in individual polls, all the results address the same issue: public mood, feelings or morale about its institutions. Also, Siegrist et al.[35] argue that while the distinction between trust and confidence is a key element of certain theories of cooperation, the dual mode approach has had little impact on empirical studies. However, Newton[40] cites Finland in 1990 where social trust remained high despite the collapse of confidence in parliament and other public institutions and Japan where low and declining levels of trust in government are accompanied by high and increasing levels of social capital. Newton's observation finds support in Luhmann's argument that a social evolution which achieves increasingly complex societies may in fact generate systems which require more confidence as a prerequisite for participation and more trust in partners as a condition for the best utilisation of chances and opportunities.

Here, there is agreement with Lipset and Schneider[26], but it is noted that for research purposes, a distinction is necessary, as argued by Luhmann[34] and extended by Siegrist et al.[35] The findings of Newton[40] are noted, that is, that cultural differences may make the distinction more important in some countries than in others as in Japan and Finland compared with Lipset and Schneider's[26] research focused on the USA. For analytical purposes, particularly in cross-cultural studies in which researchers are looking for differences and similarities, a separation of both terms is argued for here: trust for persons and confidence in institutions. As Luhmann[34] argues, confidence is best for systems and trust for partners, indicating, for example, that you can have confidence in the judiciary but you need trust for the individual judges. This study, however, relies on confidence in institutions data from the World Values Survey Wave 6.

The study is constrained by unavailability of data to compare responses to both types of question wordings (trust and confidence).

Confidence in institutions: State of research

Confidence measures have been used to monitor public perception of institutions following specific issues and events; to track changes on the longitudinal and to compare single items across countries. Price and Romantan[41] and Lipset and Schneider[26] have shown that changing confidence and its direction in one institution does not guarantee commensurate changes in another. Hoffman[42] and Twenge et al.[37] also found that there may be demographic influences. However, Cooke and Gronke[38] argue that noticeable declines in confidence may be the result of a public sceptical of many forms of power and may not necessarily be bad news.

Clausen et al.[18] found a statistically significant negative correlation between corruption and confidence in public institutions. Steen[43] compared levels of confidence in institutions across the Baltic States with France and Norway. Borowski[44] compared frequencies among the post-communist countries Lithuania, Poland, Russia, Ukraine, Czech and Buryat Republics, while Listhaug[25] examined the underlying factors influencing confidence in institutions in Norway.

Research objectives

Here frequencies (see Steen[43] and Borowski[44]) are compared and the underlying common sources of influence and distinguishing characteristics (see Listhaug[25] and Gregg and Banks[45]) are examined, specifically, in the context of unstable democracies, military influence in politics, religious beliefs, weak economies and comparably lower incomes. The following questions are asked:

1. How do the levels of confidence that African publics have in their institutions vary among the countries and compare to those of the USA?
2. What are the underlying common sources of influence that characterise the perception of institutions by Africans and how do they compare with those of the USA?

Data and methods

The data for this research were obtained from the World Values Survey Wave 6 which consisted of 17 questions, V108 to V124, listed earlier, and which are common to all countries in comparison. The question was:

> *Please look at this card and tell me, for each item, how much confidence you have in them. Is it 'A great deal' (4), 'Quite a lot' (3) 'Not very much' (2) or 'None at all' (1).*

Confidence levels were derived from the sum of 'A great deal' (4) and 'Quite a lot' (3). The sample size subject to variable ratio for Zimbabwe was 1500/17 = 88 to 1; Ghana 1552/17 = 91 to 1; Nigeria 1759/17 = 103 to 1 and the USA 2232/17 = 131 to 1. Additional World Values Survey data for Nigeria in 2000 (Wave 4), Zimbabwe 2001 (Wave 4) and Ghana 2006 (Wave 5) were also used to monitor frequency changes within countries.

Frequencies were compared, but this comparison alone is considered insufficient because of issues relating to the local meaning of constructs being influenced by culture and contextual differences.[46-48] Therefore sets of variables for shared and distinguishing characteristics were also compared. Wagner et al.[46] argue for the use of a 'cultural metric' – a set of items or variables that are inter-related and that mutually specify each other's local meaning in a culture or language group. Wagner and colleagues argue that only an interaction found in one culture that is replicated in another allows the conclusion that the effect is shared.

Cognitive and social psychology studies indicate that individuals and groups rely on several sources of meaning embedded in common sense. These multiple sources can be directly measured or observed as 'manifest' variables. The variable 'religiosity', for example, can be measured directly with the question 'how religious are you?' and the answer options on a scale of say 0 to 6. Here we expect the respondent to aggregate all the meaning sources in a single response. The same religiosity can be measured by asking several questions that address different sources of meaning for the concept, such as 'how many times do you attend religious services?', 'how many times do you pray in a day?', 'how strongly do you believe in heaven?', all also measured on a scale of 0 to 6. We can then apply a data reduction technique such as factor analysis which groups variables by similarities into sense-making groups (factors or dimensions) called 'latent' (or unobserved) variables. The result of the reduction may indicate one underlying sense-making group (unidimensional) or several groups (multidimensional). Latent variables, factors and dimensions are used interchangeably in this paper. The additional advantage of latent variables is that they make up for the inconsistencies that may be in the single-variable self-assessment.

Exploratory factor analysis was used to categorise the variables (see Appendix 1 in the supplementary material). The assumption of factor analysis is that several manifest variables depend on the same latent variable and this dependence induces a correlation (relationship) between them, denoting the existence of a common source of influence.[47,49,50] The aim of the factor analysis, as with cultural metrics, is to determine if the observed variables can be grouped into sets, thereby establishing the 'semantic scaffolding'[46] that a local notion maintains with a set of other terms.

The Factor Analysis function on the SPSS package was used for the analyses and the Promax oblique rotation was chosen over its orthogonal counterpart as there are significant correlations among the items. The extraction method was Maximum Likelihood as findings can be generalised to the larger population.[51]

Findings

Comparing frequencies

Question 1: How do the levels of confidence that African publics have in their institutions vary among the countries and compare to those of the USA?

Confidence levels were generally higher in African countries than in the USA, with the exception of the army and the police, but the rankings do vary (Table 1). Confidence levels are spread between 83% and 11% in the USA; 86% and 34% in Zimbabwe; 89% and 39% in Ghana and between 92% and 28% in Nigeria. This spread may be an indication that the countries have different approaches to scale use or that the US respondents are more sceptical[38] than their African counterparts. The highest level of confidence in the USA is in the army while in African countries, it is in the churches.

The university is the only institution that occupies one of the first four positions across all the countries. Confidence levels in the courts are quite close (55% to 57%), except for Nigeria at 46%; the rankings also vary, with US respondents ranking courts higher than the other countries. While confidence in the police is the second highest in the USA, it is the lowest in Nigeria and third to last in Ghana. Confidence in political parties occupies the bottom of the table in the USA, Zimbabwe and Ghana, but the levels are again different, and is second to last in Nigeria. Confidence in central government occupies the same position in the USA, Ghana and Zimbabwe, but is at a lower place in Nigeria. Confidence in parliament in Zimbabwe and Nigeria occupy the same position, fourth to last, and the same position in the USA and Ghana, at second to last.

In conclusion for Question 1, there are varying levels of confidence across cultures and ranking offers additional comparative statistics for single items. The higher levels of confidence in African countries do not manifestly reflect their poorer economic indices and status as emerging democracies.

Table 2 shows that confidence in churches has remained very high across the three African states, despite falling slightly in Nigeria and Ghana. Confidence in the army has also increased across all three African countries but with a higher 17-point rise in Nigeria. The last military intervention in Nigeria ended in 1999 – 1 year before the first survey in 2000; there were 12 years of uninterrupted civilian rule by the next wave in 2011. Confidence in parliament has risen marginally in Zimbabwe but dropped in Nigeria, with an even steeper drop in Ghana.

Table 1: Confidence levels ranking (%): 'A great deal' and 'Quite a lot'

USA		Zimbabwe		Ghana		Nigeria	
Army	83	Churches	86	Churches	89	Churches	92
Police	70	Universities	81	Universities	85	Banks	71
Universities	62	Charities	74	Banks	79	Television	71
Charities	62	Major companies	72	Army	79	Universities	70
Churches	59	Banks	70	Charities	76	Major companies	66
Courts	55	Environmental organisations	66	Major companies	75	Army	63
Women's organisations	51	Women's organisations	64	Women's organisations	72	Charities	62
Environmental organisations	48	Army	57	Television	72	The press	62
Civil service	46	Courts	56	Environmental organisations	69	Civil service	59
Banks	39	Civil service	53	Civil service	67	Women's organisations	57
Major companies	33	Police	52	The press	59	Environmental organisations	56
Central government	32	Central government	47	Central government	59	Unions	47
Unions	24	Television	46	Unions	58	Courts	46
Television	23	Parliament	45	Courts	57	Parliament	39
The press	22	Unions	44	Police	55	Central government	36
Parliament	19	The press	35	Parliament	54	Political parties	31
Political parties	11	Political parties	34	Political parties	39	Police	28

The percentage of 'Don't know' responses is nil for churches in Nigeria in 2000 and Ghana in 2007 and is low (1%) in Zimbabwe in 2001. It is 1–3% for the police across the countries. These figures may indicate high levels of familiarity[34,35] with the institutions and much less ambivalence in the population about their roles. For other institutions, 'Don't know' is comparatively high in Zimbabwe in 2001. The figure for unions is surprising, given their involvement in politics.[52] The year 2001 was a period of political and economic uncertainty in Zimbabwe as inflation had hit 76% and was climbing – reaching 24 400% in 2007 from 17% in 1990 when the Lancaster constitution expired and anxieties were high over land redistribution.[23,53] It is plausible, given the ongoing crisis, that the public in Zimbabwe was reluctant to express an opinion about these institutions.

Structural analysis and cultural metrics

Question 2: What are the underlying common sources of influence that characterise the perception of institutions by Africans and how do they compare with those of the USA?

Cronbach's alpha[54], a measure of internal consistency, for all 17 variables for the four countries was 0.9, which indicates high scale reliability. The Pearson's correlation table also showed significant positive correlations (association) among all items in the Ghanaian data at 0.01 level of significance and the highest was between police and courts at 0.617, while the lowest was between the unions and the churches at 0.121. All variables were significantly positively correlated at the 0.01 level for the Nigerian data, except between parliament and churches which was significant at 0.05, and the highest correlation of 0.627 was between parliament and the central government while the lowest (0.057) was between parliament and churches. All variables were significantly positively correlated for the Zimbabwean data at a 0.01 level, except environmental organisations which did not show any significance with churches (0.034) and the highest correlation of 0.663, as for the Ghanaian data, was between the police and the courts. With the US data, labour unions were not significantly correlated with churches (0.024) and environmental organisations were also not correlated significantly with churches (-0.024) and armed forces (0.026). All other variables showed significant correlations at the 0.01 level with the highest between parliament and political parties at 0.687. These results are given in the supplementary material.

A hypothetical theoretical analysis will split the 17 institutions into collectivist and individualistic countries. Another approach, informed by functions, will split them into private and public sector. Both assumptions presuppose a two-factor solution. Several solutions were considered based on the statistical outputs (scree plots), the meaningfulness of the factors and the research question. A five-factor solution was found to be most appropriate for comparison and labelled as follows: 'not-for-profit or charities'; 'political'; 'social order'; 'watchdog or fourth estate' and 'for-profit' or 'business'. The Nigerian data, however, fit best with a four-factor solution as a fifth displays only one variable which is negligible because it does not have a substantive interpretation.[50] The factors reveal the underlying order and the names are a concept operationalised by the factor. The factor labels thus comprise a set of concepts with high generalising power for cross-national studies.[45] The proportion of each factor shows its relative weight or importance in the analysis when compared with the others. The rank ordering of each variable also shows its relative importance within each meaning unit.

The figures for 'communalities', the extent to which a variable associates with others, are very low for churches in Nigeria (0.163), Ghana (0.11) and Zimbabwe (0.06), and were thus excluded. The low communalities for churches is an indication that they are quite unique and have little underlying characteristics in common with other institutions. For the USA, a communality of 0.25 for churches is also low; however, it was included in the analysis because the factor loadings were above the threshold of 0.3. Most of the other communalities[55] were above 0.40 across countries and where they dipped slightly below, the influence of the large sample sizes, the loadings on the pattern matrix and the essence of the variable's inclusion in cross-cultural comparisons were considered.

Table 2: Changing levels of confidence in African countries (%)

	Nigeria		Zimbabwe		Ghana		Nigeria	Zimbabwe	Ghana
	2000	2011	2001	2011	2007	2011	DK[a] 2000	DK[a] 2001	DK[a] 2007
Churches	95	92	82	86	90	89	0	1	0
Army	46	63	55	57	76	79	2	11	1
The press	63	62	49	35	54	59	2	11	4
Television	71	71	52	46	69	72	1	11	2
Unions	63	47	46	44	46	58	3	26	9
Police	34	28	63	52	54	55	1	3	1
Courts		46		56	62	57			1
Central government	46	36	48	47	71	59	4	7	1
Political parties	43	31	24	34		39	3	18	
Parliament	43	39	43	45	63	54	5	14	2
Civil service	69	59	53	53	59	67	3	10	4
Universities		70		81		85			
Major companies	68	66	67	72	55	75	3	12	6
Banks		71		70		79			
Environmental organisations	58	56	45	66	59	69	4	35	6
Women's organisations	50	57	46	64	64	72	7	22	5
Charities		62		74	64	76			6
N	2022	1759	1002	1500	1534	1552	2022	1002	1534

[a]*'Don't know' response*

All the statistical outputs indicate the method was appropriate. The determinant was 0.001 for the USA and Nigeria, 0.003 for Ghana and 0.002 for Zimbabwe. For all countries, the Kaiser–Meyer–Olkin Measure of Sampling Adequacy was 0.9; Bartlett's Test of Sphericity was highly significant ($p<0.001$) and the goodness-of-fit test was also highly significant ($p<0.001$). Cronbach's alpha was also computed separately for items relating to the different factors.[54,56] Minimum loading for interpretation was set at 0.3.[51,57]

The results show 10 variables with shared positions in a factor across countries – the similarities – but the positions vary. The variables are environmental NGOs, women's organisations, charities, parliament, parties, central government, television, the press, major companies and banks. There were differences in the positioning of the other seven factors denoting the unique combinations of each culture.

With the US data, the first dimension (factor) with a proportion of 34% (also referred to as explained variance) is dominant, hence most important[50] and comprises environmental organisations with the highest loading of 0.85 followed by women's organisations, charities, universities, labour unions and civil service (Table 3). The factor was named 'not-for-profit' based on the composition of the institutions and the loadings profile. Unions, however, has a cross loading on the watchdog.

The first dimension is also the most important in Ghana (35%) but with fewer institutions: women's organisations, charities and environmental organisations. The shared profile of the dominant dimension indicates similarity between the Ghanaian and US data.

The variables with shared values in the 'not-for-profit' factor for all four countries are the charities, women's organisations and environmental NGOs.

Table 3: The not-for-profit cultural metric

	USA	Zimbabwe	Nigeria	Ghana
Factor	**1**	**2**	**4**	**1**
Environmental NGOs	0.85	0.70	0.56	0.67
Women's organisations	0.82	0.74	0.91	0.82
Charities	0.70	0.63	0.66	0.80
Universities	0.58			
Labour unions	0.34*			
Civil service	0.31			
% variance	34	10	3	35
Cronbach's alpha	0.8	0.7	0.8	0.8

**Cross loading*

Dimension one for Nigeria (Table 4) is also the most important at 37% and is populated by political institutions. Central government has the highest loading of 0.87 followed by parliament and political parties. It also features the police, courts and labour unions – an association that signifies an underlying feeling of an autocratic system. It provides plausible evidence of Nigeria being more autocratic than Ghana which, despite being more collective, has been credited with the emergence of good autocrats.[32] It also

indicates that the separation of powers between the judiciary, the police and the executive branch remains in doubt in Nigeria.[9,10] The labour unions are also known to be divided between those who support the ruling elite and those against and may account for the cross loading with watchdog.

Table 4: The political cultural metric

	USA	Zimbabwe	Nigeria	Ghana
Factor	**2**	**1**	**1**	**2**
Labour unions			0.33*	
Civil service		0.39		0.38
Parliament	0.98	0.94	0.82	0.95
Political parties	0.76	0.70	0.75	0.60
Central government	0.71	0.48*	0.87	0.38*
Police			0.57	
Courts			0.53	
% variance	8	34	37	7
Cronbach's alpha	0.8	0.8	0.8	0.8

**Cross loading*

Dimension one, at 34%, is also the most important for Zimbabwe (Table 4) and is populated, as for Nigeria, by the parliament, political parties and central government, indicating strong similarity between both countries. For Zimbabwe, however, the police and courts load on the social order metric and labour unions on watchdog. Comparing Ghana and Zimbabwe, civil service loads on the political metric and central government has a cross loading with the social order category for both countries.

The rank ordering within the dimension also shows that parliament is above the parties and central government in all countries except Nigeria, where the central government is highest.

Table 5: Watchdog cultural metric

	USA	Zimbabwe	Nigeria	Ghana
Factor	**4**	**3**	**3**	**3**
Labour unions	0.33*	0.34	0.31*	
Armed forces			0.50	0.32
Television	0.85	0.77	0.66	0.80
The press	0.75	0.84	0.85	0.79
% variance	4	4	4	4
Cronbach's alpha	0.7	0.7	0.7	0.7

**Cross loading*

The press and television comprise the watchdog across countries but labour unions also load here in the USA, Nigeria (cross loadings) and Zimbabwe. The armed forces in Nigeria and Ghana also belong to this group. This is not unexpected in Nigeria and Ghana given the long period of military rule as the public may still see the soldiers as a check on the politicians. Zimbabwe had no experience of military rule as at the time of data collection in 2011. The threat of military intervention, however, remains high in African countries following the recent coup in Egypt[5] and the November 2017 putsch in Zimbabwe.

Table 6: For-profit/business cultural metric

	USA	Zimbabwe	Nigeria	Ghana
Factor	**5**	**4**	**2**	**4**
Universities		0.74	0.79	0.68
Civil service			0.47	
Major companies	0.73	0.79	0.79	0.71
Banks	0.68	0.54	0.72	0.72
Churches	0.30*			
% variance	4	3	9	3
Cronbach's alpha	0.6	0.7	0.8	0.8

**Cross loading*

The for-profit category features major companies and banks across all countries. In addition, and as a mark of different cultures, universities in Ghana, Nigeria and Zimbabwe also load strongly on this factor. The loading of universities on this metric is a major surprise as universities are traditionally for teaching and research and not profit centres. The civil service in Nigeria also loads on this factor, which may be explained by the relatively high level of corruption among public sector workers. Also, interesting for this group, is the loading of churches in the USA, although it is not very high at 0.30. Churches in the USA also load with the same level (0.30) on the social category.

Table 7: The social order cultural metric

	US	Zimbabwe	Ghana
Factor	**3**	**5**	**5**
Central government		0.45*	0.32*
Police	0.92	0.75	0.86
Courts	0.61	0.84	0.74
Armed forces	0.54	0.38	
Churches	0.30*		
% variance	6	3	3
Cronbach's alpha	0.7	0.8	0.8

**Cross loading*

The social order category also has cross loadings of central government for Zimbabwe and Ghana. The police and courts are common among three countries: the USA, Ghana and Zimbabwe. The armed forces in the USA and Zimbabwe also belong to this group in contrast with Nigeria and Ghana where they belong to the watchdog group. There are no variables for Nigeria in this category; the police and courts are on the political factor.

In conclusion for Question 2, there are five cultural metrics: not-for-profit, political, social order, watchdog and for-profit. The analyses have also shown that the underlying influences of the levels of confidence in institutions makes Ghana closer to the USA and Nigeria closer to Zimbabwe. It can also be deduced that the lack of separation of courts and police from democratic institutions in Nigeria is an indication of a continuing autocratic system and the low communalities for churches shows they are not seen by the public as an institution in the same manner in which other institutions are viewed.

Limitations and further research

The high levels of 'Don't know' in the responses for Zimbabwe need to be further interrogated to ascertain if it is a persistent situation or if the economic crisis at the time of data collection was responsible for such high levels of ambivalence.

More qualitative research through interviews and multi-country focus groups is needed to illuminate the composition of the cultural metrics across the countries. Why, for example, do Nigerians see the parliament, the political parties, the central government, police and courts as the same? Is this an indication that Nigeria is more autocratic than other African countries? Why is Ghana closer to a Western democracy, the USA, than other African countries in the sample?

Conclusions

Comparing frequencies has shown that poorer economic indices and relatively unstable democracies have not translated into comparatively lower confidence in African institutions as levels are higher than those observed in the USA. Frequencies range between 83% and 11% in the USA; 86% and 34% in Zimbabwe; 89% and 39% in Ghana and 92% and 28% in Nigeria. These percentages may, however, indicate different approaches to scale use or that US respondents are more sceptical[38] than Africans.

The rankings show that different institutions occupy different positions in the hierarchy and those with almost the same position have different levels of confidence. The army occupies the highest position in the USA and churches occupy the highest position in Africa. The university is always in the first four positions across all countries. Confidence in central government occupies the same position in the USA, Ghana and Zimbabwe, but is lower in Nigeria.

The underlying common sources of influence show that the 17 institutions can be grouped into five latent categories: 'not-for-profit'; 'political'; 'social order'; 'watchdog' and 'for-profit or business'. The positioning of the variables on the factors show the shared elements and differences which distinguish the cultures and confirms the power of the method for cross-national studies. It can be inferred that charities, women's organisations and environmental NGOs are more important and, by extension, more central to the public in the USA and Ghana, and less so in Nigeria and Zimbabwe. The political order is also more important to the public in Nigeria and Zimbabwe. This finding is crucial and adds empirical evidence to the effects of relative democratic stability in the USA and Ghana compared with Nigeria and Zimbabwe. Ghana is a surprise given that, like Nigeria and Zimbabwe, it is a young democracy with a recent history of colonial and military rule.

This study has shown that comparing frequencies and rankings, or dividing along the lines of individualistic versus collective or private and public sector, masks the dynamic distribution of the systems of meaning in the local cultures and the latent variables approach offers a more conceptually sound categorisation informed by shared and distinguishing institutions. The findings show that such measures can be used as reliable markers of culture for cross-national studies and in longitudinal studies to monitor changing perspectives.

Acknowledgements

Assistance for writing and publication was received from the MACAS-project (Mapping the Cultural Authority of Science) (www.macas-project.com) Economic and Social Research Council (ESRC), UK grant ES/K005820/1, through the Department of Psychological and Behavioural Sciences (DPBS), London School of Economics and Political Science.

References

1. Elbadawi E, Sambanis N. Why are there so many civil wars in Africa? Understanding and preventing violent conflict. J Afr Econ. 2000;9(3):244–269. https://doi.org/10.1093/jae/9.3.244
2. Gutteridge W. Undoing military coups in Africa. Third World Q. 1985;7(1):78–89. http://dx.doi.org/10.1080/01436598508419825
3. Johnson TH, Slater RO, McGowan P. Explaining African military coups d'etat, 1960-1982. Am Polit Sci Rev. 1983;78(03):622–640. http://dx.doi.org/10.2307/1961833
4. Cook SA. Ruling but not governing: The military and political development in Egypt, Algeria, and Turkey. Baltimore, MD: Johns Hopkins University Press; 2007
5. Kirkpatrick DD. Army ousts Egypt's president; Morsi is taken into military custody. New York Times. 2013 July 3. Available from: http://www.nytimes.com/2013/07/04/world/middleeast/egypt.html?_r=0
6. Olukoshi A, Mustapha AR, Soyinka W, Mnthali F. A tribute to Ken Saro-Wiwa. Rev Afr Polit Econ. 1995;22(66):471–480. http://dx.doi.org/10.1080/03056249508704152
7. French HW. Nigeria executes critic of regime; nations protest. New York Times. 1995 November 11. Available from: http://www.nytimes.com/1995/11/11/world/nigeria-executes-critic-of-regime-nations-protest.html?pagewanted=all
8. Reuters. Parcel bomb kills Nigerian journalist accused of plotting. New York Times. 1986 October 20. Available from: http://www.nytimes.com/1986/10/20/world/parcel-bomb-kills-nigerian-journalist-accused-of-plotting.html
9. Idowu W. Corruption, the police and the challenges of a free and fair election in Nigeria. JSDA. 2010;12(7):1–12.
10. Oko O. Seeking justice in transitional societies: An analysis of the problems and failures of the judiciary in Nigeria. Brook J Int'l L. 2005;31(1):9–82.
11. Kallaway P. Apartheid and education: The education of black South Africans. Johannesburg: Ravan Press; 1984.
12. Wolpe H. Capitalism and cheap labour-power in South Africa: From segregation to apartheid. Econ Soc. 1972;1(4):425–456. http://dx.doi.org/10.1080/03085147200000023
13. Martin D, Johnson P, Mugabe R. The struggle for Zimbabwe: The Chimurenga war. London: Faber & Faber; 1981.
14. Ellis S. The ANC in exile. Afr Affairs. 1991;90(360):439–447.
15. Hofstede G. Culture's consequences: Comparing values, behaviors, institutions and organizations across nations. 2nd ed. Thousand Oaks, CA: Sage; 2001.
16. Cohen AB, Wu MS, Miller J. Religion and culture: Individualism and collectivism in the East and West. J Cross-Cult Psychol. 2016;47(9):1236–1249. https://doi.org/10.1177/0022022116667895
17. Cohen AB. Religion's profound influences on psychology morality, intergroup relations, self-construal, and enculturation. Curr Dir Psychol Sci. 2015;24(1):77–82. https://doi.org/10.1177/0963721414553265
18. Clausen B, Kraay A, Nyiri Z. Corruption and confidence in public institutions: Evidence from a global survey. The World Bank Econ Rev. 2011;25(2):212–249. http://dx.doi.org/10.1596/1813-9450-5157
19. Transparency International. Corruption Perception Index for year 2011. Berlin: Transparency International; 2011. Available from: https://www.transparency.org/cpi2011/results
20. World Bank. GDP per capita (current US$). World Bank national accounts data, and OECD National Accounts data files. Washington DC: World Bank; 2016. Available from: http://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2012&start=1960
21. Clarida R, Gali J, Gertler M. Monetary policy rules and macroeconomic stability: Evidence and some theory. Q J Econ. 2000;115(1):147–180. https://doi.org/10.1162/003355300554692

22. Svensson LE. Inflation forecast targeting: Implementing and monitoring inflation targets. Eur Econ Rev. 1997;41(6):1111–1146. https://doi.org/10.1016/S0014-2921(96)00055-4

23. World Bank. Inflation, consumer prices (annual %). International Monetary Fund, International Financial Statistics and data files. Washington DC: World Bank; 2016. Available from: http://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?end=2012&start=1960

24. Newton K, Norris P. Confidence in public institutions: Faith, culture or performance? In: Pharr SJ, Putnam RD, editors. Disaffected democracies: What's troubling the trilateral countries. Princeton, NJ: Princeton University Press; 1999. p. 52–73.

25. Listhaug O. Confidence in institutions: Findings from the Norwegian Values Study. Acta Sociol. 1984;27(2):111–122. https://doi.org/10.1177/000169938402700202

26. Lipset SM, Schneider W. The confidence gap: business, labour and government in the public mind (Rev ed.). Baltimore, MD: Johns Hopkins University Press;1987.

27. Triandis HC. The psychological measurement of cultural syndromes. Am Psychol. 1996;31(4):407–415. http://dx.doi.org/10.1037/0003-066X.51.4.407

28. Shweder R, LeVine RA. Culture theory. Essays on mind, self, and emotion. Cambridge, UK: Cambridge University Press; 1984.

29. Gartstein MA, Carranza JA, González-Salinas C, Ato E, Galián MD, Erickson NL, et al. Cross-cultural comparisons of infant fear: A multi-method study in Spain and the United States. J Cross-Cult Psychol. 2016;47(9):1178–1193. https://doi.org/10.1177/0022022116663851

30. Mascolo MF. The coactive construction of selves in cultures. New Dir Child Adolesc Dev. 2004;2004(104):79–90. http://dx.doi.org/10.1002/cd.105

31. Triandis HC. Individualism and collectivism. Boulder, CO: Westview Press; 1995.

32. Gorodnichenko Y, Roland G. Culture, institutions and democratization. NBER Working Paper no. 21117. JEL no. H1,P48,Z1. Cambridge, MA: National Bureau of Economic Research; 2015. http://dx.doi.org/10.3386/w21117

33. De Raad B, Morales-Vives F, Barelds DP, Van Oudenhoven JP, Renner W, Timmerman ME. Values in a cross-cultural triangle: A comparison of value taxonomies in the Netherlands, Austria, and Spain. J Cross-Cult Psychol. 2016;47(8):1053–1075. https://doi.org/10.1177/0022022116659698

34. Luhmann N. Familiarity, confidence and trust: Problems and alternatives. In Gambetta D, editor. Trust: Making and breaking of cooperative relations. Oxford: Blackwell; 1998. p. 94–107.

35. Siegrist M, Gutscher H, Earle TC. Perception of risk: The influence of general trust, and general confidence. J Risk Res. 2005;8(2):145–156.

36. Tyler TR. Public trust and confidence in legal authorities: What do majority and minority group members want from the law and legal institutions. Behav Sci Law. 2001;19(2):215–235. http://dx.doi.org/10.1002/bsl.438

37. Twenge JM, Campbell WK, Carter NT. Declines in trust in others and confidence in institutions among American adults and late adolescents 1972–2012. Psychol Sci. 2014;25(10):1914–1923. http://dx.doi.org/10.1177/0956797614545133

38. Cook TE, Gronke P. The sceptical American: Revisiting the meanings of trust in government and confidence in institutions. J Polit. 2005;67(3):784–803. http://dx.doi.org/10.1111/j.1468-2508.2005.00339.x

39. Hager MA, Hedberg EC. Institutional trust, sector confidence, and charitable giving. J Nonprofit Public Sec Mark. 2016;28(2):164–184. http://dx.doi.org/10.1080/10495142.2015.1011508

40. Newton K. Trust, social capital, civil society, and democracy. Int Polit Sci Rev. 2001;22(2):201–214. https://doi.org/10.1177/0192512101222004

41. Price V, Romantan A. Confidence in institutions before, during, and after 'Indecision 2000'. J Polit. 2004;66(3):939–956. https://doi.org/10.1111/j.1468-2508.2004.00284.x

42. Hoffman JP. Confidence in religious institutions and secularization: Trends and implications. Rev Rel Res. 1998;39(4):321–343. https://doi.org/10.2307/3512442

43. Steen A. Confidence in institutions in post-communist societies: The case of the Baltic States. Scand Polit Stud. 1996;19(3):205–225. https://doi.org/10.1111/j.1467-9477.1996.tb00391.x

44. Borowski A. Confidence in social institutions in the post-communist countries. ILSHS. 2014;25:7–17. https://doi.org/10.18052/www.scipress.com/ILSHS.25.7

45. Gregg PM, Banks AS. Dimensions of political systems: Factor analysis of a cross-polity survey. Am Polit Sci Rev. 1965;59(3):602–614. https://doi.org/10.2307/1953171

46. Wagner W, Hansen K, Kronberger N. Quantitative and qualitative research across cultures and languages: Cultural metrics and their application. Integr Psychol Behav. 2014;48(4):418–434. https://doi.org/10.1007/s12124-014-9269-z

47. Treier S, Jackman S. Democracy as a latent variable. Am J Polit Sci. 2008;52(1):201–217. https://doi.org/10.1111/j.1540-5907.2007.00308.x

48. Pardo R, Calvo F. Attitudes towards science among the European public: A methodological analysis. Public Underst Sci. 2002;11(2):155–195. https://doi.org/10.1088/0963-6625/11/2/305

49. Bartholomew DJ, Steele F, Galbraith J, Moustaki I. Analysis of multivariate social science data. Boca Raton, FL: CRC Press; 2008.

50. Marks G, Steenbergen MR. European integration and political conflict. Cambridge, UK: Cambridge University Press; 2004. http://dx.doi.org/10.1017/CBO9780511492013

51. Fields A. Discovering statistics using SPSS. 2nd ed. London: Sage; 2005.

52. Raftopoulos B. The labour movement and the emergence of opposition politics in Zimbabwe. Labour Cap Soc. 2000;33(2):256–286.

53. Moyo S. The political economy of land acquisition and redistribution in Zimbabwe, 1990-1999. J S Afr Stud. 2000;26(1):5–28. http://dx.doi.org/10.1080/030570700108351

54. Cronbach LJ. Coefficient alpha and the internal structure of tests, Psychometrika. 1951;16(3):297–334. http://dx.doi.org/10.1007/BF02310555

55. Velicer WF, Fava JL. Effects of variable and subject sampling on factor pattern recovery. Psychol Methods. 1998;3(2):231–251. http://dx.doi.org/10.1037//1082-989X.3.2.231

56. Cortina JM. What is coefficient alpha? An examination of theory and application. J Appl Psychol. 1993;78(1):98–104. http://dx.doi.org/10.1037/0021-9010.78.1.98

57. Tabachnick BG, Fidell LS. Using multivariate statistics. Boston, MA: Allyn and Bacon; 2001.

Household food waste disposal in South Africa: A case study of Johannesburg and Ekurhuleni

AUTHORS:
Suzan Oelofse[1]
Aubrey Muswema[1]
Fhumulani Ramukhwatho[1]

AFFILIATION:
[1]Natural Resources and the Environment, Council for Scientific and Industrial Research, Pretoria, South Africa

CORRESPONDENCE TO:
Suzan Oelofse

EMAIL:
soelofse@csir.co.za

DATES:
Received: 18 Aug. 2017
Revised: 23 Oct. 2017
Accepted: 07 Dec. 2017
Published: 30 May 2018

KEYWORDS:
food security; municipal solid waste; socio-economic status; developing country; sub-Saharan Africa

HOW TO CITE:
Oelofse S, Muswema A, Ramukhwatho F. Household food waste disposal in South Africa: A case study of Johannesburg and Ekurhuleni. S Afr J Sci. 2018;114(5/6), Art. #2017-0284, 6 pages. http://dx.doi.org/10.17159/sajs.2018/20170284

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Gauteng Infrastructure Financing Agency (South Africa)

Food waste is becoming an important issue in light of population growth and global food security concerns. However, data on food wastage are limited, especially for developing countries. Global estimates suggest that households in developed countries waste more food than those in developing countries, but these estimates are based on assumptions that have not been tested. We therefore set out to present primary data relating to household food waste disposal for South Africa within the sub-Saharan African context. As the Gauteng Province contributes about 45% of the total municipal waste generated in South Africa, the case study area covers two of the large urban metropolitan municipalities in Gauteng, namely Ekurhuleni and Johannesburg, with a combined population of 8.33 million, representing nearly 15% of the South African population. Municipal solid waste characterisation studies using bulk sampling with randomised grab sub-sampling were undertaken over a 6-week period during summer in 2014 (Johannesburg) and 2016 (Ekurhuleni), covering a representative sample of the municipal waste collection routes from households in each of the two surveyed municipalities. The food waste component of the household waste (excluding garden waste) was 3% in Ekurhuleni and 7% in Johannesburg. The results indicate that an average of 0.48 kg (Ekurhuleni) and 0.69 kg (Johannesburg) of food waste (including inedible parts) is disposed of into the municipal bin per household per week in the two municipalities, respectively. This translates into per capita food waste disposal of 8 kg and 12 kg per annum, respectively, in South Africa as compared to the estimated 6–11 kg per annum in sub-Saharan Africa and South and Southeast Asia.

Significance:

- Research on food waste in developing countries is limited.
- This study is the first of its kind undertaken in South Africa.
- Food waste research is important to address food security issues.
- This study provides evidence to support Sustainable Development Goal 12.3.

Introduction

Global estimates suggest that the world population will reach 9.6 billion by 2050.[1] At current lifestyle patterns, it is estimated that we will need the equivalent of nearly three planets to sustain the world population in 2050.[1] Increasing food production to ensure food security will put additional pressure on already constrained natural resources.[1] Yet, it is estimated that between a third and a half of all food produced globally never reaches the point of human consumption.[2] It is ethically unacceptable to waste food that could be used to feed people.[3] In South Africa, 12 million people (24.5% of the national population) go to bed hungry each day[4] and it is reported that South Africa has the largest proportion of food wastage in Africa[5]. Food waste can be seen as having a triple negative impact: (1) the waste of resources (including water and energy) used along the supply chain in the production, handling and distribution of food that is not consumed by humans; (2) the socio-economic impacts associated with food insecurity; and (3) environmental impacts associated with waste and emissions (including greenhouse gas emissions) generated during the production, harvesting, processing, distribution and disposal of food that is not consumed.[6]

All households, irrespective of income level, contribute to food waste.[7-9] Baker et al.[10] found a direct relationship between household food waste generation, household income and the number of people residing in a household. Food waste decreases when there are more occupants, yet increases with household income level.[10] As such, households in wealthier countries tend to generate more food waste than those in developing countries; while in all cases there are significant economic costs associated with food waste. The average annual cost of household food waste in South Africa is ZAR21.7 billion (EUR1.5 billion at an exchange rate of ZAR14.40/EUR1.00)[11]. Annual costs of food waste in Europe are reported as ZAR7 200 million (EUR500 million) in Finland[12], ZAR388.8 billion (EUR27 billion) in Norway[13], ZAR30.96 billion (EUR2.15 billion) in Denmark[14] and ZAR126.7 billion in the UK (GBP10.2 billion at an exchange rate of ZAR12.42/GBP1.00)[15]. Therefore, preventing food wastage will not only save money for households, but will have broader economic, social and environmental benefits.[16] Reducing food waste will address food and water security concerns[17] and contribute to the development of more sustainable food systems[18].

The choice of food waste disposal system by households can have environmental implications (e.g. composting is preferable to throwing food in the trash), but the largest environmental benefits will be derived from preventing food waste altogether.[19] Research found that 60% of the total household food waste in the UK is avoidable.[20] Initial findings of ongoing research in South Africa noted similarities in the food types being wasted as well as the reasons for wasting food between households in the City of Tshwane and the UK.[21] The main reasons why food is thrown away in the UK is because it either 'wasn't used in time', or too much was cooked, prepared or served.[20] Similar

reasons for food wastage are reported for households in South Africa.[9,21] The potential therefore exists that food waste at household level in South Africa can be reduced.

Comparing per capita food waste at consumer level between different studies is problematic because of non-standard definitions and sampling methods when measuring food waste.[22] Attempts have therefore been made to standardise the definition of food waste. For example, a project commissioned by the European Union proposes the following definition:

> *Any food, and inedible parts of food, removed from the food supply chain to be recovered or disposed (including composted, crops ploughed in/not harvested, anaerobic digestion, bio-energy production, co-generation, incineration, disposal to sewer, landfill or discarded to sea).*[23]

The data reported in this paper include the edible and inedible parts of food that is disposed of by households into the municipal bin but exclude food waste being composted, fed to animals or disposed of in any manner other than in the municipal bin.

Current estimates of food wastage (edible portion only) throughout the value chain suggest that 'on a per capita basis, much more food is wasted in the industrialised world than in developing countries'[24]. The water and carbon footprint of the food wasted in the industrial world extends beyond the boundaries of the country in which the food is wasted, to the countries where the food is produced and processed.[17] The environmental impacts of wasted food are therefore a global issue in light of globalised food markets.

The per capita food wasted (edible portion) by consumers in Europe and North America is estimated to be 95 kg and 115 kg per annum, respectively; as compared to only 6 kg and 11 kg per annum in sub-Saharan Africa and South/Southeast Asia, respectively.[24] If we assume that food wastage patterns in South Africa are similar to those in the rest of sub-Saharan Africa, then South African households contribute about 4% of total food waste in South Africa.[6] The latest published estimates of the magnitude of food waste at consumer level in South Africa, using estimates from Gustavsson et al.[24] regarding the proportion of food wasted at each stage of the value chain (on average for sub-Saharan Africa), together with the Food and Agriculture Organization of the United Nations data on actual food production in South Africa, is 501 000 tonnes per annum[25]. This translates into 9.68 kg per capita per annum using 2011 population statistics for South Africa.[26]

In an earlier study, following a different methodology (i.e. using estimates from waste characterisation studies conducted in South Africa across various municipalities rather than Gustavsson et al.'s averages for sub-Saharan Africa), Nahman et al.[11] estimated household food wastage in South Africa at 1.4 million tonnes per annum, or 28 kg per capita per annum. This is nearly three times the 9.68 kg per capita per annum estimated by Nahman and De Lange[25]. However, Nahman et al.[11] caution that the 28 kg per capita per annum may be an overestimation, because of the difficulties of extrapolating food waste quantities from the reported organic waste quantities derived from waste composition analysis.

A study of households in the City of Tshwane, South Africa, found the average food waste generation to be 6 kg per household per week across income groups (with an average three people per household),[27] which is equivalent to about 98 kg per capita per annum. In this study, source separated food waste from 133 participating households was collected and weighed.[27] The study by Ramukhwatho[27] provides an accurate measure of the actual food waste generation across income groups prior to treatment or disposal, whereas the report by Nahman et al.[11] estimated the average food waste disposed of in the municipal bin.

The results from the studies by Nahman et al.[11] and Ramukhwatho[27] suggest that post-consumer food waste in South Africa is higher than in the rest of sub-Saharan Africa. South Africa is one of the high-income countries in the region, with a larger and more advanced economy in comparison with the rest of sub-Saharan Africa.[28] Higher household incomes are associated with a decline in consumption of starchy food staples and increased diversification of diets towards more fresh fruit and vegetables, dairy, meat and fish.[22] However, the study by Ramukhwatho[27] found that starchy foods (mealie pap, bread and rice) are still the most wasted food types across income groups with higher-income groups wasting more rice and vegetables as compared to middle- and low-income households. Wealthier people are reported to buy more food as a result of affordability[29] and are therefore likely to waste more. However, Koivupuro et al.[30] found no correlation between households' income levels and the amount of food wasted in Finland. Household food wastage in urban South Africa can be expected to be higher than in the rest of sub-Saharan Africa as a result of affordability, but this expectation still needs to be confirmed.

The small sample size (133 households) included in the study by Ramukhwatho[27], and the difficulties expressed in the paper by Nahman et al.[11], highlight the need for more accurate data on household food waste from a representative sample of households in South Africa. The data reported in this paper were collected as part of municipal solid waste characterisation studies in Johannesburg and Ekurhuleni – two of the larger urban metropolitan municipalities in the Gauteng Province of South Africa during the summers of 2014 and 2016. A limitation of reporting the results of two independent studies is that only directly comparable data can be used and, therefore, although both characterisation studies covered at least two seasons, only data collected during the summer seasons are reported here.

Contextualising the study area

Gauteng is the smallest of the nine provinces of South Africa, comprising only 1.5% of the land area, yet it is home to 25% of the national population[26] and contributes 34% of the country's gross domestic product (GDP)[31]. The Gauteng Province consists of three metropolitan municipalities – Ekurhuleni, Johannesburg and Tshwane – as well as two district municipalities – Sedibeng and West Rand – which are further subdivided into six local municipalities (Figure 1). Johannesburg is home to 36% of Gauteng's population, Ekurhuleni 26%, Tshwane 24% and the two district municipalities 7% each.[32] Being the most densely populated and urbanised province of South Africa (97% of the population in Gauteng is urbanised), it is estimated that Gauteng contributes about 45% of total municipal solid waste generated in South Africa.[33] Credible data on waste generation and composition in the rural areas of Gauteng are limited.

Household waste collection in the urban areas of Gauteng is typically done once a week in line with the National Domestic Waste Collection Standards.[34] Waste separation at source (S@S) is not yet common practice in South African households. Following the promulgation of the *National Environmental Management: Waste Act, 2008*[35] and approval of the National Waste Management Strategy, which sets targets for, among other things, initiating S@S programmes for households in all metropolitan municipalities, secondary cities and large towns[36] – a number of such initiatives have been launched in selected suburbs in the study area. S@S initiatives primarily target recyclables (i.e. paper, cardboard, plastics, glass and metals), and therefore food waste is still collected as part of the mixed household solid waste collection service provided by municipalities or their service providers, and disposed of at landfill.

Methodology

Sampling

Waste characterisation studies in Johannesburg and Ekurhuleni were undertaken using bulk sampling with randomised grab sub-sampling, as described by Jarrod Ball and Associates Consortium[37], covering a representative sample of the municipal waste collection routes from households in each of the two surveyed municipalities. This sampling methodology was used in order to collect data that are comparable to the 2001 characterisation results for Johannesburg as reported by Jarrod Ball and Associates Consortium[37]. The sampling areas were chosen purposefully to cover a range of socio-economic areas (high, middle and low income), but the routes within the selected areas were randomly chosen.

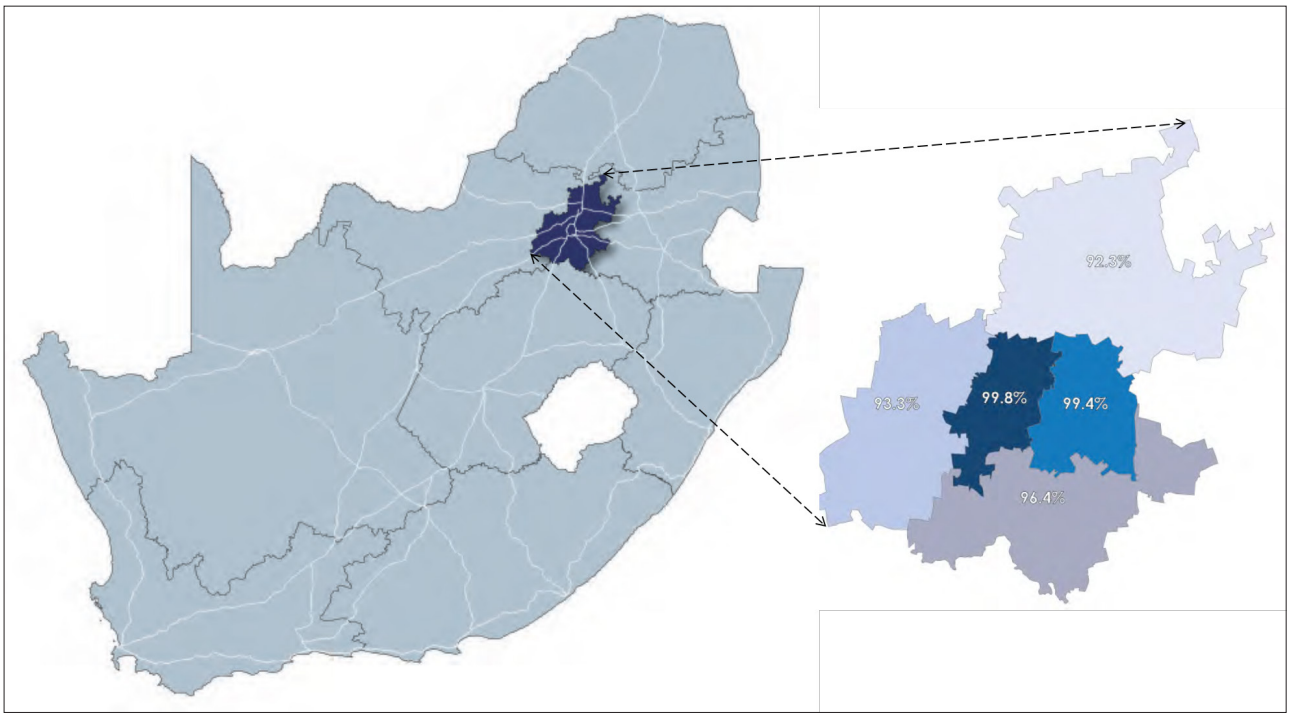


Figure 1: Geographical location of Gauteng Province, South Africa; and the percentage urbanised population per municipal area (adapted from Gauteng Provincial Government[31]).

Sampling teams consisted of one person accompanying the waste collection vehicle during normal waste collection from households to record data on the route and the number of households serviced, and another four people to collect and sort the waste samples at the landfill.

Bulk samples were collected using municipal waste collection vehicles following the normal collection schedules (weekly collection). At the end of the collection routes, the mass of the waste load collected was determined using the weighbridge at the landfill. A random grab sample of 100–200 kg was then taken from each load and manually sorted into pre-determined waste categories, including food waste, by trained waste sorters. Sorting was done at dedicated sorting areas on the landfill sites, away from the active workface (the landfill operational area). Any food and inedible parts of food (including drinks and cooking oil) that were present in the waste sample were sorted into the food waste category (if it was still in packaging material, the packaging was removed). Each sorted waste fraction was weighed using a calibrated scale and accurately recorded. At the end of each day, the sorted waste was returned to the workface for proper disposal.

Calculations

Waste disposal per household (kilograms/household/week) was calculated by dividing the mass of each bulk sample by the number of households serviced by the specific vehicle. Then, the mass of food waste per household for each sample was calculated by multiplying the percentage food waste determined from each sub-sample with the total calculated household waste for the bulk sample from which the sub-sample was taken.

Limitations

It was not in all instances possible to clearly determine the number of households from which waste was collected in the case of townhouse complexes, apartment buildings and communal collection points. Therefore, the results reported here only include data collected from routes for which the number of households serviced could clearly be determined; data from townhouse complexes, gated communities and apartment buildings are not included in the results presented here.

Using bulk sampling of waste collected by the municipal waste collection service, limits the results to the food waste disposed into the municipal bin. Food and drink waste fed to pets, disposed of into the sink and treated through home composting is not included in the scope of this study.

Results and discussion

The data reported were collected from 74 waste collection routes covering 44 927 households in the Johannesburg metropolitan municipality and 41 routes covering 20 439 households in the Ekurhuleni metropolitan municipality (Table 1). Data in both municipalities were collected in summer, which is the rainy season in Gauteng Province.

Table 1: Samples obtained and analysed

Samples	Johannesburg	Ekurhuleni
Total number of collection routes sampled	74	41
Total number of households included in the samples	44 927	20 439
Combined weight of bulk samples collected (kg)	938 220	80 380
Combined weight of sub-samples analysed (kg)	10 761	13 829
Average weight of sub-samples (kg)	145.42	246.72
Percentage of bulk sample analysed	1.15%	17.2%

The average food waste disposed per household in Johannesburg (0.69 kg/week) is higher than that in Ekurhuleni (0.48 kg/week) across all income categories (Table 2).

The differences in weekly food waste disposal (kg/hh) between income groups and municipalities appear to be small and therefore a statistical analysis was done to determine the significance of the differences between high, middle and low income as well as between high and low income. The Mann–Whitney test was done because of the skewness of the data. The results indicate that there are no statistically significant differences in the food waste generation (kg/hh), i.e. there is a consistent pattern of differences between municipalities and income levels (Table 3).

Table 2: Food waste disposed of by high-, middle- and low-income households into the municipal bin (kilograms/household/week) in two metropolitan cities in Gauteng

Income category	Average (kg/hh)	Median (kg/hh)	Minimum (kg/hh)	Maximum (kg/hh)	Number of bulk samples	Number of households sampled
Johannesburg metropolitan municipality						
High	0.72	0.56	0.00	3.00	29	15 719
Middle	0.56	0.01	0.00	2.92	12	5967
Low	0.73	0.34	0.00	4.66	33	23 241
All	0.69	0.32	0.00	4.66	**74**	**44 927**
Ekurhuleni metropolitan municipality						
High	0.53	0.19	0.00	1.67	19	8599
Middle	0.45	0.20	0.00	1.76	18	9357
Low	0.38	0.26	0.05	0.96	4	2483
All	0.48	0.19	0.00	1.76	**41**	**20 439**

Table 3: Mann–Whitney test results of comparison of food waste between income groups and the two municipalities ($p > 0.05$ implies no statistically significant difference)

Comparison	*p*-value	Interpretation
Comparing municipalities		
Ekurhuleni to Johannesburg	0.604	Not significant
Comparing Ekurhuleni to Johannesburg within income categories		
High	0.5839	Not significant
Middle	0.2009	Not significant
Low	0.7149	Not significant
Comparing income categories		
High-middle-low	0.2131	Not significant
High-low	0.8730	Not significant
Middle-low	0.1173	Not significant
High-middle	0.1240	Not significant
Comparing income categories within municipalities		
High-middle-low in Johannesburg	0.2294	Not significant
High-middle-low in Ekurhuleni	0.9371	Not significant
High-low in Johannesburg	0.6817	Not significant
High-low in Ekurhuleni	1.0000	Not significant
Middle-low in Johannesburg	0.0958	Not significant
Middle-low in Ekurhuleni	0.9664	Not significant
High-middle in Johannesburg	0.1604	Not significant
High-middle in Ekurhuleni	0.7035	Not significant

It is interesting to note that 27% of all samples collected in Johannesburg and 14% collected in Ekurhuleni did not contain any food waste. This is very odd as all households are assumed to generate at least some unavoidable food waste in the form of egg shells, bones etc. However, this may be as a result of the random grab sampling method used for sub-sampling or it may be an indication that some households use alternative food waste management strategies which may include feeding of food waste to pets and/or home composting.

Global estimates of food waste assume similar food wastage in South Africa compared with the rest of sub-Saharan Africa.[24] For comparative purposes, the data collected in this study from metropolitan households in South Africa were converted into food waste per capita per annum. This results in estimates of approximately 8 kg and 12 kg per capita per annum in Ekurhuleni and Johannesburg, respectively (Table 4); which is higher than the average of 6 kg per capita per annum for sub-Saharan Africa, but lower than the 95 kg in Europe.[24] Furthermore, the estimates for sub-Saharan Africa and Europe include all food that was meant for human consumption and that reached the consumer, but was never consumed[24]; whereas our results include only the waste disposed of into the municipal bin.

Table 4: Calculated per capita food waste disposed of per annum using 2011 census data from StatsSA[26]

Variable	Johannesburg	Ekurhuleni
Population	4 439 991	3 178 470
Number of households	1 437 168	1 015 645
Per cent urbanisation	99.80	99.40
Number of urban households	1 434 294	1 009 551
People per household	3	3
Food waste (average kilograms/household/week)	0.69	0.48
Food waste (kilograms/capita/week)	0.22	0.15
Food waste (kilograms/capita/annum)	11.61	7.98
Food waste disposed of by urban households (tonnes/week)	989.66	484.58
Food waste disposed of by urban households (tonnes/annum)	51 462.46	25 198.40

The study by Ramukhwato[27] was conducted in urban households in Gauteng which kept food diaries; 86% of respondents mostly wasted starchy staples (i.e. mealie pap, rice and bread), and 17% managed their food waste either by feeding it to pets (14%) or by home composting (3%). The results reported here are therefore likely to be a conservative estimate of household food waste generation in Gauteng.

Based on the 2011 urban population of each municipality, the amount of food waste disposed to landfill by urban households in Ekurhuleni and Johannesburg can be calculated at between 25 198 tonnes and 51 462 tonnes per annum.

Conclusions and recommendations

An average of 0.48 kg and 0.69 kg of food waste (including inedible parts) is disposed of into the municipal bin per household per week in the municipalities of Ekurhuleni and Johannesburg, respectively. These figures translate into per capita food waste disposal of about 8 kg and 12 kg per annum, respectively, and between 25 198 tonnes and 51 462 tonnes from households per annum.

Because the results reflect only the waste disposed of into the municipal bin, and it is reported that other food waste management strategies are also employed by households in Gauteng[27], this should be seen as a conservative estimate for urban households. It is important to note that rural households were not included in the survey, and therefore extrapolation beyond urban households in Gauteng using these data is not possible.

It is evident from this study that initiatives focusing on urban households' food waste reduction and diversion from landfill have the potential to divert significant amounts of food waste from landfill. The results from this study will specifically be useful to the two surveyed municipalities, as they provide accurate data against which the success of food waste reduction strategies can be measured.

It is recommended that future research should consider different sampling methods in addition to bulk sampling, including random sampling of individual household bins or bags across a city or by asking participants to keep food diaries and record the reasons for the food wastage at household level. All future research should aim to collect complementary evidence for the development of food waste reduction strategies for South Africa including data across the different types of municipalities in South Africa (A, B1, B2, B3 and B4) to obtain representative data across metros, large cities and more rural towns.

Acknowledgements

We acknowledge the efforts and hard work of the sampling and sorting teams who assisted with data collection under sometimes very trying conditions. The cooperation and kind assistance of the management, staff and service providers of the two sampled municipalities are acknowledged as well as Renee Koen for assistance with the statistical analysis and Anton Nahman for valuable input into an earlier draft of the paper. The research was funded by the Gauteng Infrastructure Financing Agency and we are grateful for their support and the permission of all involved to publish these findings.

Authors' contributions

S.O. conceptualised the article, identified relevant data sets, extracted relevant data, interpreted the data and drafted the bulk of the article. A.M. was involved in the data collection, contributed to the methodology section and assisted with the interpretation of the data and conclusions. F.R. assisted with the literature review and drafting of the introduction.

References

1. United Nations. Responsible consumption and production: Why it matters [document on the Internet]. c2015 [cited 2016 Nov 21]. Available from: http://www.un.org/sustainabledevelopment/wp-content/uploads/2016/08/16-00055L_Why-it-Matters_Goal-12_Consumption_2p.pdf
2. Institution of Mechanical Engineers. Global food, waste not want not. c2013 [cited 2013 Jan 03]. Available from: https://www.imeche.org/policy-and-press/reports/detail/global-food-waste-not-want-not
3. Thyberg KI, Tonjes DJ. Drivers of food waste and their implications for sustainable policy development. Resour Conserv Recycl. 2016;106:110–123. https://doi.org/10.1016/j.resconrec.2015.11.016
4. Hosken G. Twelve million going to bed hungry in SA. TimesLIVE. 2013 January 30 [cited 2013 Feb 06]. Available from: http://www.timeslive.co.za/thetimes/2013/01/30/twelve-million-going-to-bed-hungry-in-sa
5. AllAfrica. South Africa: Waste not, want not is a lesson still to be learnt. Business Day. 2010 September 20 [cited 2014 May 13]. Available from: http://allafrica.com/stories/201009201320.html
6. Oelofse SHH, Nahman A. Estimating the magnitude of food waste generated in South Africa. Waste Manag. Res. 2013;31(1):80–86. https://doi.org/10.1177/0734242X12457117
7. Sustainability Victoria. Food waste avoidance studies 2010 [document on the Internet]. c2011 [cited 2017 Jan 26]. Available from: http://www.google.co.za/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=0ahUKEwjz1rHFqN_RAhXkAcAKHRNlAvYQFggXMAA&url=http%3A%2F%2Fwww.sustainability.vic.gov.au%2F-%2Fmedia%2Fresources%2Fdocuments%2Fpublications-and-research%2Fpublications%2Fc--f%2Fpublications-food-waste-avoidance-studies-2010.pdf%3Fla%3Den&usg=AFQjCNGbZgTDKWTZHSa1MzvpO_JkDPcjuA&sig2=vrl_5GZCsQZwckpRHlZLEg
8. Visagie J. Growth in the middle class: Two perspectives that matter for policy. Develop South Afr. 2015;32(1):3–24. https://doi.org/10.1080/0376835X.2014.965387
9. Ramukhwatho FR, Du Plessis R, Oelofse S. Household food wastage by income level: A case study of five areas in the city of Tshwane Metropolitan Municipality, Gauteng Province, South Africa In: Proceedings of WasteCon 2016; 2016 October 17–21; Johannesburg, South Africa. Johannesburg: IWMSA; 2016. p. 57–64.
10. Baker D, Fear J, Dennis R. What a waste – An analysis of household expenditure on food [document on the Internet]. c2009 [cited 2014 Mar 08]. Available from: http://www.tai.org.au/node/1580%20food
11. Nahman A, De Lange W, Oelofse S, Godfrey L. The costs of household food waste in South Africa. Waste Manage. 2012;32(11):2147–2153. https://doi.org/10.1016/j.wasman.2012.04.012
12. Statistics Finland. Annual national accounts [webpage on the Internet]. c2011 [cited 2013 Nov 14]. Available from: http://www.stat.fi/til/vtp/index_en.html
13. Olufssøn HK. Prevention of food waste. Presentation of the ForMat-project [document on the Internet]. c2012 [cited 2013 Oct 09]. Available from: http://ostfoldforskning.no/uploads/dokumenter/Food%20Waste%20juni%202012/Prevention%20of%20food%20waste%20%20Halfdan%20Kverneland%20Olafss%C3%B8n,%20Matvett%20AS.pdf
14. Danish Agriculture and Food Council. Food waste: A global tragedy. Aarhus: Aarhus University, CONCITO, the Institute of Food and Resource Economics; 2011. p. 31.
15. WRAP (Waste Resource and Action Programme). Food waste report: The food we waste [document on the Internet]. c2008 [cited 2011 May 26]. Available from: http://wrap.s3.amazonaws.com/the-food-we-waste.pdf
16. Schott A, Canovas A. Current practice, challenges and potential methodological improvements in environmental evaluations of food waste prevention. Resour Conserv Recyl. 2015;101:132–142. https://doi.org/10.1016/j.resconrec.2015.05.004
17. WRAP, WWF-UK. The water and carbon footprint of household food and drink waste in the UK: Final report 2011 [document on the Internet]. c2011 [cited 2017 Jan 26]. Available from: http://waterfootprint.org/media/downloads/Water-and-carbon-footprint-food-and-drink-waste-UK-2011.pdf
18. Quested TE, Parry AD, Easteal S, Swannell R. Food and drink waste from households in the UK. Nutr Bull. 2011;36:460–467. https://doi.org/10.1111/j.1467-3010.2011.01924.x
19. Quested TE, Marsh E, Stunell D, Parry AD. Spaghetti soup: The complex world of food waste behaviours. Resour Conserv Recyl. 2013;79:43–51. https://doi.org/10.1016/j.resconrec.2013.04.011
20. WRAP (Waste Resource and Action Programme). Household food and drink waste in the United Kingdom 2012: Final report [document on the Internet]. c2013 [cited 2015 Jan 08]. Available from: http://www.wrap.org.uk/sites/files/wrap/hhfdw-2012-main.pdf.pdf

21. Oelofse SHH, Marx-Pienaar N. Household food wastage – A case study of middle to high income urban households in the City of Tshwane. In: Proceedings of WasteCon 2016; 2016 October 17–21; Johannesburg, South Africa. Johannesburg: IWMSA; 2016. p. 51–56.

22. Parfit J, Barthel M, Macnaughton S. Food waste within food supply chains: Quantification and potential for change to 2050. Philos Trans R Soc Lond B Biol Sci. 2010;365:3065–3081. https://doi.org/10.1098/rstb.2010.0126

23. Fusions. Reducing food waste through social innovation: Fusions definitional framework for food waste. Full report 3 July 2014. Göteborg: Swedish Institute for Food and Biotechnology; 2014.

24. Gustavsson J, Cederberg C, Sonesson U, Van Otterdijk R, Meybeck A. Global food losses and food waste: Extent, causes and prevention. Study conducted for the International Congress SAVE FOOD! at Interpack 2011, Düsseldorf, Germany. Rome: Food and Agriculture Organisation of the United Nations; 2011.

25. Nahman A, De Lange W. Costs of food waste along the value chain: Evidence from South Africa. Waste Manage. 2013;33:2493–2500. https://doi.org/10.1016/j.wasman.2013.07.012

26. Statistics South Africa (StatsSA). Census 2011: Census in brief. Report no. 03-01-41. Pretoria: StatsSA; 2012.

27. Ramukhwatho FR. An assessment of the household food wastage in a developing country: A case study of five areas in the City of Tshwane Metropolitan Municipality, Gauteng Province, South Africa [MSc dissertation]. Pretoria: University of South Africa; 2016. http://uir.unisa.ac.za/handle/10500/21162

28. International Monetary Fund (IMF). World economic and financial surveys: Regional economic outlook: Sub-Saharan Africa. Navigating Headwinds [document on the Internet]. c2015 [cited 2017 Jan 11]. Available from: https://www.imf.org/external/pubs/ft/reo/2015/afr/eng/pdf/sreo0415.pdf

29. Skourides I, Smith SR, Loizides M. Sources and factors controlling the disposal of biodegradable municipal solid waste in urban and rural areas of Cyprus. Waste Manag Res. 2008;26:188–195. https://doi.org/10.1177/0734242X07085756

30. Koivupuro HK, Hartikainen H, Silvennoinen K, Katajajuuri JM, Heikintalo N, Reinikainen A, et al. Influence of socio-demographical, behavioural and attitudinal factors on the amount of avoidable food waste generated in Finnish households. Int J Consum Stud. 2012;36(2):183–191. https://doi.org/10.1111/j.1470-6431.2011.01080.x

31. Gauteng Provincial Government. Gauteng spatial development framework 2030. Johannesburg: Gauteng Provincial Government; 2016 [cited 2017 Jan 10]. Available from: http://www.gautengonline.gov.za/Documents/GSDF%202030.pdf

32. Statistics South Africa (StatsSA). Census 2011 municipal report – Gauteng. Report no. 03-01-5572. Pretoria: StatsSA; 2012.

33. Department of Environmental Affairs (DEA). National waste information baseline report. Pretoria: DEA; 2012.

34. Republic of South Africa. National domestic waste collection standards. GN 21, GG 33935 of 21 January 2011. Pretoria: Government Printers; 2011.

35. Republic of South Africa. National environmental management: Waste act, 2008 (Act No 59 of 2008). GG 32000, Vol. 525, Notice no. 278. Pretoria: Government Printers; 2009.

36. Department of Environmental Affairs (DEA). National waste management strategy. GN 344 GG 35306 of 4 May 2012. Pretoria: DEA; 2012.

37. Jarrod Ball and Associates Consortium. A waste stream analysis of the general waste stream. DANCED Environmental Capacity Building Project, Johannesburg. Mini-Project WM4 on Metro-wide waste management planning – Phase 1. Current waste management in the City of Johannesburg. Report no. 1A; December 2001. Unpublished report.

Revelations from an online diagnostic arithmetic and algebra quiz for incoming students

AUTHORS:
Aneshkumar Maharaj[1]
Thokozani Dlomo[1]

AFFILIATION:
[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

CORRESPONDENCE TO:
Aneshkumar Maharaj

EMAIL:
maharaja32@ukzn.ac.za

DATES:
Received: 16 May 2017
Revised: 24 Oct. 2017
Accepted: 11 Dec. 2017
Published: 30 May 2018

KEYWORDS:
preparedness; diagnostics; pre-calculus; performance; examinations

HOW TO CITE:
Maharaj A, Dlomo T. Revelations from an online diagnostic arithmetic and algebra quiz for incoming students. S Afr J Sci. 2018;114(5/6), Art. #2017-0078, 7 pages. http://dx.doi.org/10.17159/sajs.2018/20170078

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Eskom; National Research Foundation: South African Agency for Science and Technology Advancement

A review of relevant literature revealed that research indicates an alarming number of secondary school students who enter tertiary institutions display a serious lack of mathematical basics required for the study of university mathematics. This so-called change from the secondary school to tertiary study of mathematics has become an important field of mathematical learning research. In this context, at the University of KwaZulu-Natal, online diagnostics for pre-calculus knowledge and skills were made available to first-year mathematics students. We report here on some findings from an analysis of the provision of diagnostic questions for basic arithmetic and algebra. The results show that the diagnostic questions which were in the form of an online quiz assisted students to cope with their mathematics module in the sense that it positively impacted on the final results of students who attempted the quiz.

Significance:

- Students who completed an online quiz on basic arithmetic and algebra to identify their strengths and weaknesses indicated that the quiz prepared them for the material they had to study for their mathematics module.

Introduction

The increase in the number of students entering university has revealed a greater inconsistency in the background levels of those students enrolled for first-year mathematics modules.[1] This transition from secondary to tertiary education in mathematics creates a complex phenomenon covering a huge range of problems and issues.[2] Although it is believed that there is evidence of similar transitional issues in other disciplines in science, it seems that the transition in mathematics is by far the most serious and the most problematic.[3,4] A possible reason for this discrepancy is the many changes that take place in the transition from secondary to tertiary education. These include changes in teaching and learning styles, the type of mathematics that is taught, conceptual understanding, procedural knowledge that is required to advance through and the changes in the level of advanced mathematical thinking needed.[5] News24, a local online newspaper, reported that, according to Umalusi (a South African council for quality assurance in general and further education and training), matric (Grade 12) mathematics results were worse in 2014 than in the previous years.[6] This news report supports the perception that there is a very serious decline in core mathematical skills. Because a percentage of those matric students enrol for tertiary studies in mathematics in subsequent years, this issue should be of concern to those who lecture to such students, especially in first-year mathematics modules.

Jennings[2] is of the opinion that it has now become even more important to understand a student's prior knowledge of mathematics before they study tertiary mathematics. The main reasons for this need are: (1) to find out whether a student would benefit from additional support on basic arithmetic and algebra, offered in an online format and (2) to produce suitable first-year students who are going to be competent throughout their course of study. These are not the only reasons, but we have chosen in this paper to focus on these two given reasons. For students to be competent throughout their course of study, necessary steps need be taken to improve their performance, which implies that all possible assistance should be available to students to reach the ultimate goal of succeeding in their mathematics course. For this reason, at the University of KwaZulu-Natal (UKZN), the provision of online diagnostic questions for basic mathematical knowledge and skills for incoming students was rolled out at the beginning of February 2015. Keeping in mind the fact that not all the students attempted the online quiz, for this paper the final performance of students who completed the quiz was considered for the analysis and compared with the performance of those students who did not take the quiz.

Research question

The main research question was: What does an investigation into completion of the online diagnostic basic arithmetic and algebra quiz by incoming university students reveal? To answer this question the following sub-questions were formulated for the provision that was in place for 2015: (1) What was the participation level of students with regard to the online diagnostic quiz? (2) What was the performance level of students with regard to the online diagnostic quiz? (3) How did students' participation in the quiz affect their performance in the final examination? (4) What are the views of students on the impact of this online diagnostic quiz?

Literature review

There has been a huge concern about the inflow of students to tertiary level institutions who are lacking the fundamentals in mathematics. It is perceived that the majority of students have difficulties with mathematical basics in arithmetic and algebra. Breiteig and Grevholm[7] noted that the abstractness of algebra is one reason for students' problems. This problem impacts on the study of calculus because algebra is a pre-requisite for such studies. This lack of student preparedness has consequences for student and staff satisfaction, student self-esteem, retention and progression of students, and financial costs to universities and to students arising through failure to progress or complete a course.[3] According to Kaput[8], algebra is a route to higher mathematics, but at the same time a

barrier for many students which forces them to take another educational direction. On the transition from arithmetic to algebra, Moseley and Brenner[9] noted that ' …moving from arithmetic problems to working with more symbolic representations of relationships with variables in algebra is an area of persistent difficulty for many' (p. 4). The provision of diagnostic questions for basic arithmetic and algebra focused on in this paper was intended to assist the students to develop the ability to see variables as representations of unknown quantities that, despite their abstract nature, can still be altered by the arithmetic operations.[9] Looking at the challenges that students have concerning mathematics and also considering the importance of mathematics in our modern society, it is not surprising that learning support in mathematics is becoming ever more important.[10]

To address the under-preparedness of incoming students to study mathematics, many universities offer mathematical support services such as tutors, online resources and individual consultations in order to improve the results of the students.[11] The majority of institutions in Ireland and internationally now make use of diagnostic testing for incoming first-year students.[11] This kind of support gives students an opportunity to revise some basic concepts in mathematics. It is believed that, generally, diagnostic testing has helped to indicate the level of preparedness of students who are entering tertiary level and has also contributed to some level of improvement in their final results.[11] Thus it bridges the gap between the level of preparedness expected or required upon entry level and the mathematical capabilities acquired by new students. At UKZN, such online support was put in place. Students attempt the quizzes online to improve their knowledge or to practise a particular topic. Online quizzes can be attempted an unlimited number of times and there is no time limit to complete the quiz questions. Thus, a student has the opportunity to work on the quiz questions and improve on their basic knowledge and skills considered relevant for the study of calculus. The article by Maharaj and Wagh[12] provides detailed information on how the quiz questions were formulated based on the identified expected learning outcomes for basic arithmetic and algebra. Those identified outcomes were used to frame suitable online diagnostic questions. In this paper, we focus on the provision of those diagnostics to students.

Conceptual framework

Our conceptual framework was informed by the literature review and was guided by the following:

1. The study of mathematics is hierarchical in nature.[13]
2. To study university calculus, students need to have adequate knowledge and skills relating to basic arithmetic and algebra.[4]
3. The provision of suitable online diagnostic questions on basic arithmetic and algebra could improve the level of student competencies with regard to these.[11,12]
4. There was a need to research the impact of the provision of online diagnostics for pre-calculus as offered at UKZN before a decision could be arrived at as to whether the quizzes should be made compulsory for all incoming students who want to study first-year calculus.

Methodology and participants

The UKZN learning MOODLE site for Diagnostics for Pre-calculus was used for data collection. On this site, students had access to five pre-calculus quizzes that were designed to focus on relevant knowledge and skills for the study of calculus. The students had to first complete an online consent letter, as required by the research office at UKZN for ethical clearance (protocol reference number HSS/1058/014CA). The Basic Arithmetic and Algebra quiz was used to gather data on the participants who completed the quiz. The quiz consisted of 24 questions which involved purely basic mathematical knowledge in arithmetic and algebra. Each question carried a maximum of 2 marks. The system awarded 2 marks if a student submitted the correct answer for their first attempt. If the first attempt was incorrect, the system awarded 1 mark for a correct second or third attempt. A final mark of zero was awarded for a question if the first three or more attempts were incorrect. For the study, only those students who completed all the quiz questions were regarded as having taken the quiz. Students who submitted their responses to only some of the questions were considered as having not attempted the quiz.

The quiz was attempted by students from different first-semester mathematics modules offered at UKZN: Introduction to Calculus (Math130), Mathematics for Natural Sciences (Math150), Mathematics 1A Engineering (Math131) and Quantitative Methods (Math134). In total, there were 1888 attempts for the quiz, which indicated the overall number of attempts. That number was reduced to 997 based on the criteria that only students who completed the quiz were considered for the study. For the purpose of this study, it was decided to investigate the impact of the quiz on the final results of students enrolled for the Math130 and Math150 modules. The reason for this choice was that Math130 is a core module for students who want to continue with further studies in mathematics, while Math150 is a compulsory module for other students who want to study science.

A total of 250 students enrolled for Math130, and out of those, 95 students attempted the quiz, so 155 students did not attempt the quiz. A total of 872 students enrolled for Math150, and out of those, 392 students attempted the quiz, so 480 did not attempt the quiz. For the students who attempted the quiz more than once, only their highest-graded complete attempt was considered. Students' responses for the quiz were used to gather information to answer the research question. On the system, each question was allocated a maximum of 2 points. As indicated earlier, if the first attempt submitted was incorrect then there was a penalty for subsequent attempts. After studying the students' performances, it was decided to use the following interval mark ranges for each question: 1.70 and below; 1.70 to 1.89; 1.90 and above.

A questionnaire was also formulated to obtain students' views on the diagnostic quiz and how it could be improved. Participants were chosen based on the criteria that they used the learning site to attempt the quiz questions. An email was sent to such students indicating that research was being conducted to gauge the effectiveness of the provision of the online diagnostics in which they participated and that they were invited to complete the online questionnaire. Also, the final examination results for all the students registered for the two modules were obtained from the UKZN Student Management System website. Those final results were used to determine if there were differences in the final scores between the students who attempted the quiz and those who did not.

The final examination results lists were obtained for Math130 and Math150. Each list was used to separate the students into two groups: students who completed the quiz and students who did not complete the quiz. Because we were comparing the two population means within a large data set, the two-sample z-statistical test was used, which follows a normal distribution and the population samples are independent.[14] The *Z*-test formula used was:

$$Z_{calc} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where Z_{calc} is a calculated Z value; and $\bar{x}$, s, μ and n represent the sample mean, standard deviation, population mean and the number of students who attempted the quiz, respectively, for both groups. Setting up the hypothesis test, we wanted to determine if there is a difference in the final examination results of the students who attempted the quiz compared to those who did not. Thus our null hypothesis was: attempting the quiz had no effect on final examination results. That is:

$H_0 : \mu_1 - \mu_2 = 0$ or $H_0 : \mu_1 = \mu_2$

Against the alternative hypothesis:

$H_1 : \mu_1 > \mu_2$

According to the null hypothesis, our population mean difference is zero. In the formula, this is represented as $(\mu_1 - \mu_2)_0$.

Not all students wrote the final examinations. Possible reasons for not writing the final examination include that students dropped out of the module or did not obtain Duly Performed (DP) Certificates. The data for such students were excluded from the study. For the testing of this hypothesis, only student attempts for Math130 and Math150 were considered. As indicated before, the former is a mainstream module undertaken by students who want to continue with studies in mathematics, while the latter is a compulsory service module offered to science students.

Entry requirements for the mathematics modules

The following entry requirements were obtained from the UKZN College of Agriculture, Engineering and Science handbook.[15] For Math130 (Introduction to Calculus), the pre-requisite requirement was: Higher Grade D or Standard Grade A for matric Mathematics, or NSC Level 5 Maths. For Math150 (Mathematics for Natural Sciences), the pre-requisite requirement was: Higher Grade E or Standard Grade B for matric or NSC Level 4 Maths. The minimum pre-requisite requirements for each of those two modules are regarded as being equivalent by the College of Agriculture, Engineering and Science. For each of those two modules, the students who were enrolled satisfied the minimum requirements. In a case in which a student did not meet the minimum requirements, the student was required to enrol for a foundation mathematics module, called Math199. After a student passed Math199 with a specified minimum percentage, then they qualified to enrol for Math130 or Math150. The purpose of Math199 is to equip students with mathematical tools that are needed for their chosen mathematics module. Table 1 indicates the percentage ranges for the different levels and their meaning, which were in operation from 2008 onwards for the National Senior Certificate examinations (NSC). Because some participants completed their Grade 12 examinations before the year 2008, during which period symbols were used, the necessary conversions were done to match the old Grade 12 symbols with the new levels.

Table 1: Grade 12 proficiency levels and their meanings

Level	Percentage	Competence
7	80–100	Outstanding achievement
6	70–79	Meritorious achievement
5	60–69	Substantial achievement
4	50–59	Adequate achievement
3	40–49	Moderate achievement
2	30–39	Elementary achievement
1	0–29	No achievement

Participants' Grade 12 mathematics and English proficiency levels

In this section we focus on the mathematics and English proficiencies of the Math130 and Math150 students as determined by the NSC examinations. For each module, bar graphs represent visually the spread of the NSC levels and the frequency for each of those levels.

Figure 1 indicates that the majority of students in both the groups had proficiency levels from 5 to 7, with regard to their senior certificate mathematics and English results. Also note that from the Math130 entry requirements it was required that students who wanted to be enrolled for Math130 had to have obtained at least a level 5 in their senior certificate mathematics. There was also a small number of students who achieved less than level 5. Those students were the ones who did foundation mathematics (Math199), and after passing with at least 60% were enrolled for Math130.

Figure 2 indicates that the majority of students attained levels from 4 to 6 with regard to their results for senior certificate mathematics and English. Note that for Math150, the entry requirement was a pass at at least a level 4 for mathematics. There was also a small number of students who achieved less than level 4. Those students were the ones who did foundation mathematics (Math199), and after passing with at least 50% were enrolled for Math150.

With the above focus on the Grade 12 entry requirements and their frequencies, it was revealed that for each module the two student groups considered were not of different ability levels with regard to their school achievement for mathematics and English. For mathematics, they could be grouped into one category because of the entry requirements for each module. The implication here is that in testing the hypothesis for each module, the participants who completed and those who did not complete the quiz could be regarded as being of similar ability levels with regard to their mathematics and English proficiency when they were enrolled for their university mathematics modules.

Analysis, findings and discussion

The analyses and findings are presented under the following sub-headings: Average student scores per question; student participation and performance levels; and student views on impact of quiz.

Average student scores per question

These results are for all the students across the different modules who attempted and completed the quiz on Basic Arithmetic and Algebra.

Looking at Figure 3, which gives the average student scores for each question, we observe that the questions could be placed into one of three categories, for which the students: (1) scored an average of 1.70 and below, (2) between 1.70 to 1.89 and (3) an average of 1.90 and above. A high average shows that most students scored full marks on the particular question, that is, they correctly answered the question. Likewise, when the average is low it shows that on that question students did not perform well.

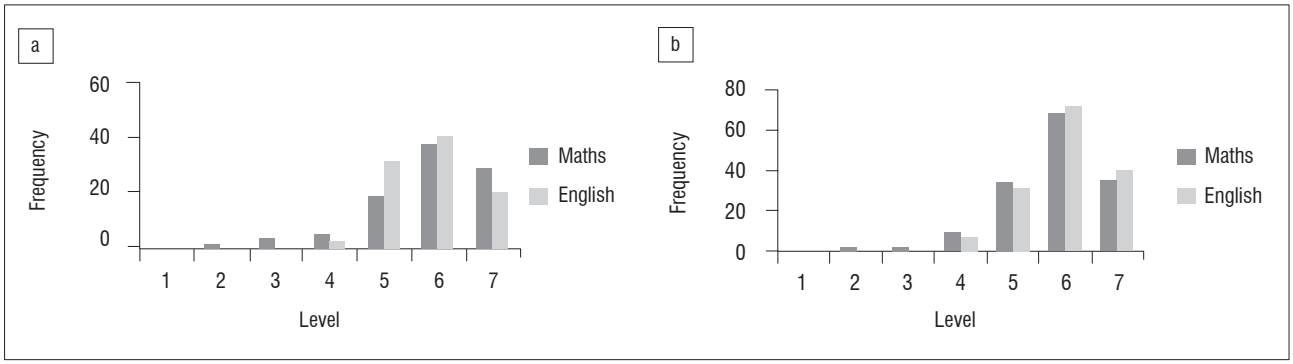


Figure 1: NQF Mathematics and English proficiency levels for Math130 students who (a) completed the quiz and (b) did not complete the quiz.

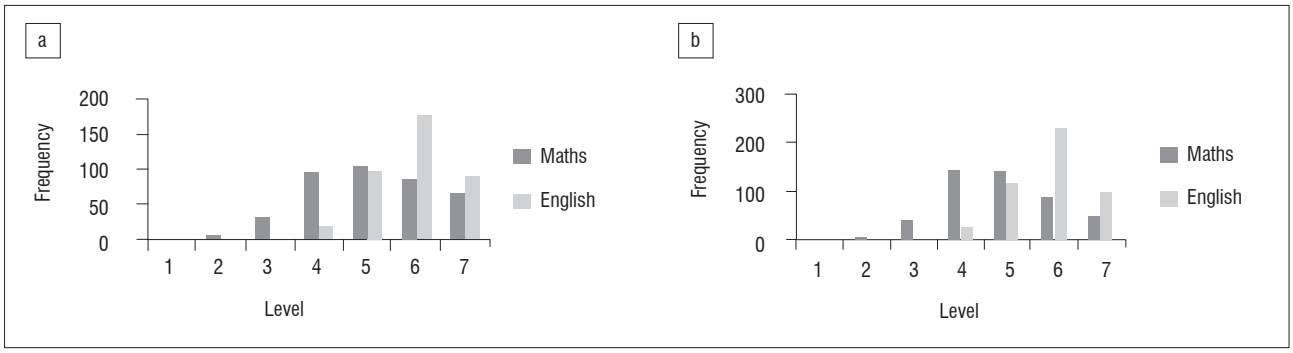


Figure 2: NQF Mathematics and English proficiency levels for Math150 students who (a) completed the quiz and (b) did not complete the quiz.

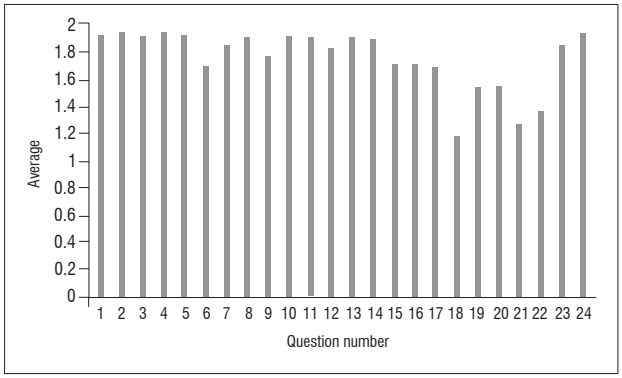


Figure 3: Average student scores for each question (n=997).

Scores of 1.70 and below

Looking at Figure 3, student performance in Questions 17 to 22 was below the average of 1.70. This score indicates that students had difficulty in answering these questions. The following are the types of questions for which students scored 1.70 and below:

$\log_{\sqrt{2}} 2 = \ldots$; $\log_a 2 + \log_a 5 = \log_a \ldots$; $\log_a 2 - \log_a 5 = \log_a \ldots$; $7\log_2 a = \log_2 \ldots$;
$\log 2 + \log 5 = \ldots$; $\log 16 \div \log 8 = \ldots$

The finding is that questions on the application of basic properties for logarithms were not adequately understood by students. The implication is that first-year mathematics lecturers should be aware of and plan to address this shortcoming. If one accepts the principle of moving from the known to the unknown, then it is important to determine and address this shortcoming before the students are exposed to the natural logarithm and its notation, that is ln$x$.

Scores of 1.70 to 1.89

As can be seen in Figure 3, students performed in this intermediate range for Questions 6, 7, 9, 12, 15, 16 and 23. The following are the types of questions for which students scored marks in this interval:

$\frac{1}{x} + \frac{1}{x^2} = \ldots$; $\frac{1}{x} - \frac{1}{x^2} = \ldots$; $\frac{1}{x} \div \frac{1}{x^2} = \ldots$; $2x^3 + 3x^3 = \ldots$; $(2x^3)^2 = \ldots$; $(2x^3)^2 = \ldots$;
$(2x^{x+1})^3 = \ldots$; If $2x + 5 = a$ then $x = \ldots$

The finding here is that, although students performed better, these types of questions could become stumbling blocks to their success when faced with calculus-related questions that require knowledge and skills based on the following: (1) addition, subtraction and division of algebraic fractions, (2) adding like terms, (3) applying the raising a power to a power law for exponents and (4) solving a linear equation in which a constant is represented by a letter. The implication is that measures need to be put in place to improve the proficiency of students with regard to these fundamentals.

Scores of 1.90 and above

It is evident from Figure 3 that average student performance was the highest for Questions 1–5, 8, 10, 11, 13, 14 and 24. This gives the impression that these questions were simple and easy to answer for the participants. The following are the types of questions for which students scored in this interval range:

$\frac{1}{2} + \frac{1}{3} = \ldots$; $\frac{1}{2} - \frac{1}{3} = \ldots$; $\frac{1}{2} \times \frac{1}{3} = \ldots$; $\frac{1}{2} \div \frac{1}{3} = \ldots$; $\frac{1}{2} \times \frac{1}{2} = \ldots$; $\frac{1}{x} \times \frac{1}{x^2} = \ldots$;
$\frac{1}{x} \times \frac{1}{x} = \ldots$; $2x^3 \times 3x^5 = \ldots$; $2x^3 - 3x^3 = \ldots$; $2x^3 \div (3x^5) = \ldots$; if $2x + 5 \geq 3$, then $x \geq \ldots$

The finding here is that students displayed adequate proficiency for knowledge and skills based on: (1) addition, subtraction, multiplication and division of arithmetic fractions, (2) multiplication of simple algebraic fractions of the type indicated, (3) applying the multiplication and division laws of exponents for which base is the same, (4) subtracting like terms and (5) solving simple linear inequalities. The average student scores suggest that students were more proficient in subtracting like terms than they were at adding like terms, which seems strange; the question on the subtraction of like terms seems to be more difficult because it requires the calculation 2 – 3 compared with the easier calculation of 2 + 3.

Student participation and performance levels

These results are presented under the following sub-headings: Students enrolled for Math130 and students enrolled for Math150.

Table 2 indicates that of 250 students who wrote the Math130 examinations, 95 students completed the quiz and 155 did not. For Math150, 872 students wrote the exam, 392 completed the quiz and 480 did not. We note that average examination marks for both modules were higher for students who completed the quiz compared with students who did not. For both modules, there was a large number of students who did not attempt the quiz, because participation was voluntary.

Table 2: Final examination mean percentages of students who attempted the quiz and those who did not

Module	Attempted quiz	Did not attempt
Introduction to Calculus (Math130)	n = 95 mean = 63.94%	n = 155 mean = 58.01%
Mathematics for Natural Sciences (Math150)	n = 392 mean = 67.68%	n = 480 mean = 63.73%

Students enrolled for Math130

Here we present our findings on the impact of the quiz on the average final examination marks of students who completed the quiz (n=95) and those who did not (n=155).

Figure 4 shows the overall performance in the Math130 final examinations for students who completed the quiz and those who did not. From Figure 4 and Table 2, we observe that students who completed the quiz scored an average examination mark of 63.94% and those who did not scored an average mark of 58.01%. So the students who completed the quiz scored on average about 6% higher in their final examination compared with students who did not take the quiz. The implication here seems to be that completing the quiz impacted positively on the final examination results of the students. However, before any conclusions can be reached, we need to determine if the difference is statistically significant.

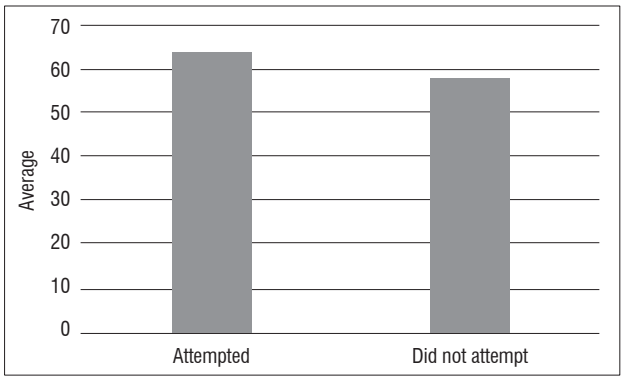


Figure 4: Average exam percentages for Math130 students who attempted and completed the quiz and those who did not.

A 99% level of significance ($\alpha=0.01$) was used. We tested the null hypothesis:

$H_0 : \mu_1 - \mu_2 = 0.$

Versus the alternative hypothesis:

$H_1 : \mu_1 > \mu_2 .$

The test statistic $Z_{calc}=3.128$ and H_0 is rejected if $Z_\alpha < Z_{calc}$ where Z_α was obtained from a Normal Z table. As $Z_\alpha=Z_{0.01}$ and $Z_{calc}=3.128$, it follows that $Z_\alpha < Z_{calc}$ so we conclude that H_0 is rejected. Thus, the results verify that the difference is statistically significant at a 99% confidence level, that is, that the students who completed the quiz performed better in their examinations than those who did not.

Students enrolled for Math150

We present our findings on the impact of the quiz on the average final examination marks of students who completed the quiz ($n=392$) and those who did not ($n=480$).

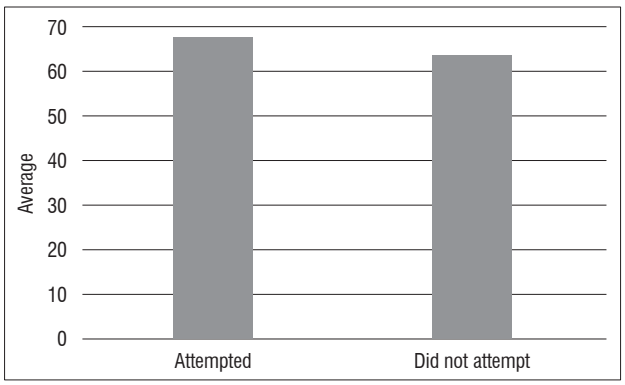


Figure 5: Average final exam percentages for Math150 students who attempted and completed the quiz and those who did not.

Figure 5 illustrates the overall average performance of Math150 students who completed the quiz and those who did not. From Figure 5 and Table 2, we observe that students who completed the quiz scored an average final exam mark of 67.68% and those who did not scored 63.73%. So students who took the quiz scored on average about 4% higher in their final examination than students who did not take the quiz. The implication seems to be that attempting the quiz improved the final examination results of the students. However, once again, before any conclusions can be reached, we need to determine whether the difference is statistically significant.

A 99% level of significance ($\alpha=0.01$) was used. We tested the null hypothesis:

$H_0 : \mu_1 - \mu_2 = 0.$

Versus the alternative hypothesis:

$H_1 : \mu_1 > \mu_2 .$

The test statistic arrived at was $Z_{calc}=4.142$. Z_α was obtained from a Normal Z table. As $Z_\alpha=Z_{0.01}=2.58$ and $Z_{calc}=4.142$, the implication is that $Z_\alpha < Z_{calc}$ so we conclude that H_0 should be rejected. Thus the results verify that the difference is statistically significant at a 99% confidence level, that is, that the students who completed the quiz performed on average better in their final examinations than those who did not.

Student views on impact of quiz

A questionnaire consisting of 10 multiple-choice and open-ended questions was completed by 60 students from both Math130 and Math150. The questions and student responses are indicated in Table 3.

Table 3 indicates that the majority of the respondents felt that the online diagnostic quiz was *helpful* to them, it was *easy* to access the diagnostic quiz online and that they were *likely* to continue attempting diagnostic quizzes online. Furthermore, students felt that the online diagnostic quiz related *well* to the course topics and that the online diagnostic quiz on Basic Arithmetic and Algebra helped them to understand key concepts in the relevant university mathematics module for which they were registered. This response is quite significant because the diagnostic quiz focused on identified knowledge and skills for basic arithmetic and algebra which are prerequisites for the modules for which the respondents were enrolled. The response to Question 6 indicates that the majority of the respondents were satisfied with the online diagnostic quiz. Note that at least 15% indicated that they were extremely satisfied. This response is worth noting as extremely satisfied was not indicated as a possible option in the framing of the question. The finding from the data was that the majority of the respondents rated the importance of attempting the online quiz towards the higher end of the scale, that is 7 or more. The responses for Questions 5 to 9 indicate that the majority of the students rated highly that taking the quiz was of benefit to them. The second and third responses that are indicated as common responses to Question 10 could be used to improve on the online diagnostic quiz. It is clear that some students want more challenging quiz items. Also, for students who do not answer correctly, there is a need to give more feedback that could help them. Perhaps the idea of building in pop-ups to further assist such students, as was suggested by a previous study[16], should be considered.

Conclusions and recommendations

Although participation in the quiz was voluntary, it was found that a satisfactory number of students registered for each of the modules on which the study focused completed the quiz. For the Introduction to Calculus (Math130) module, the participation from registered students was 38% while for the Mathematics for the Natural Sciences (Math150) module it was 45%. Quiz items that were based on the application of basic laws for logarithms gave the students the most difficulty. Further, it was found that the students who completed the quiz performed better in their examinations than those who did not attempt or complete the quiz. This was determined at a 99% confidence level for each of the two modules, Math130 and Math150. For those modules it was found that the majority of students in both groups (those who completed and those who did not complete the quiz) had similar proficiency levels with regard to their senior certificate mathematics and English results.

Table 3: Summary of questions and responses from students ($n=60$)

	Question	Student response									
1	Did you attempt the online quiz on Basic Arithmetic and Algebra?	Out of the 60 responses, 35 indicated ' yes' and 25 indicated 'no' .									
2	What were the reasons for you not attempting the quiz?	Responses for not attempting the quiz included: • 'I did not find the time to do it.' • 'I wasn't confident about my answers or if I still understood the concept, so decided not to attempt it at all.' • 'I had back to back tests and online assignments that were due in a short period of time.' • 'Each time I try to attempt this quiz an error always occurs. It was not compulsory.' • 'I did not have time & I felt that it would not benefit me.' • 'I couldn't find it on moodle.'									
		Students' responses in percentages (%)									
		Extremely	**Quite**	**Moderately**	**Slightly**	**Not at all**					
3	How ***helpful*** is the online diagnostic quiz?	18.6	46.5	25.6	7.0	2.3					
4	How ***easy*** is it to access the diagnostic quiz online?	36.6	36.6	12.2	7.3	7.3					
5	How ***likely*** are you to continue attempting the diagnostic quizzes online?	23.3	23.3	20.9	16.3	16.3					
6	Overall, are you satisfied with the online diagnostic quiz, neither satisfied nor dissatisfied with it, or dissatisfied with it?	15.0	55.0	25.0	0	5.0					
7	How ***well*** did the online diagnostic quiz relate to the course topics?	35.0	35.0	17.5	12.5	0					
8	How ***much*** did the online diagnostic quiz help you in understanding the key concepts in your maths course?	19.5	31.7	36.6	7.3	4.9					
9	On a scale from 1 to 10, how would you rate the importance of attempting the online quiz?	**Scale percentage responses**									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
		7	0	5	5	7	7	16	16	16	21
10	Do you have any other comments about how the online diagnostic quiz could be improved?	The following were some of the common responses: • 'It's all perfect.' • 'Add more interesting questions or challenging ones.' • 'If you get an incorrect answer let there be an explanation of why this answer is wrong and what might [have] driven you to choose it.' • 'The online quiz is okay, the problem is that I don't get enough time to attempt it because I spend most of my time doing tutorials.'									

Thus the quiz should be made compulsory for incoming students who want to study calculus, because the research shows that there was a positive correlation between completing the quiz and final examination results. This positive correlation was for both the modules. Also from the questionnaire responses it was noted that the majority of those students who responded felt that the online diagnostic quiz benefitted them and contributed to their understanding of key concepts in the mathematics module for which they were enrolled. These responses suggest that those students appreciated that the basic concepts and skills for arithmetic and algebra formed the foundation for their university calculus modules.

We recommend that the quiz should be made compulsory because the research shows that it had a positive impact on students' final examination results. It is also recommended that the lecturing staff should analyse the results of quiz items to determine which pre-calculus knowledge and skills require attention and address these areas so that subsequent teaching of the relevant calculus concepts is likely to be beneficial to their students.

Acknowledgements

We acknowledge the National Research Foundation (South Africa) for supporting T.D. to be part of their internship programme for 2015/2016 and for making available grants for the project 'Online diagnostics for undergraduate mathematics'. We thank UKZN for allowing us to use the data on MOODLE and also the students' results for research purposes. Eskom is acknowledged for providing grants for the UKZN–Eskom Mathematics Project, which laid the groundwork for assisting students to cope with their university mathematics studies.

Authors' contributions

A.M. was responsible for the conceptualisation of the study and proofreading the paper. T.D undertook the data collection. Both authors analysed the data and wrote the paper.

References

1. Hoyles C, Newman K, Noss R. Changing patterns of transition from school to university mathematics. Int J Math Educ Sci Technol. 2001;32(6):829–845. https://doi.org/10.1080/00207390110067635
2. Jennings M. Issues in bridging between senior secondary and first year university mathematics. In: Hunter R, Bicknell B, Burgess T, editors. Crossing divides: Proceedings of the 32nd Annual Conference of the Mathematics Education Research Group of Australasia. Volume 1. Palmerston North, NZ: MERGA; 2009. p. 273–280.
3. Croft AC, Harrison MC, Robinson CL. Recruitment and retention of students – An integrated and holistic vision of mathematics support. Int J Math Educ Sci Technol. 2009;40(1):109–125. https://doi.org/10.1080/00207390802542395
4. Kajander A, Lovric M. Transition from secondary to tertiary mathematics: McMaster University experience. Int J Math Educ Sci Technol. 2005;36:149–116. https://doi.org/10.1080/00207340412317040
5. Hong YY, Kerr S, Klymchuk S, McHardy J, Murphy P, Thomas M, et al. Teacher's perspective on the transition from secondary to tertiary mathematics education. In: Hunter R, Bicknell B, Burgess T, editors. Crossing divides: Proceedings of the 32nd Annual Conference of the Mathematics Education Research Group of Australasia. Volume 1. Palmerston North, NZ: MERGA; 2009. p. 241–248. https://doi.org/10.1080/00207390903223754
6. Maths, physics results worse than 2013 – Umalusi. News24. 2014 December 30. Available from: http://www.news24.com/SouthAfrica/News/Maths-physics-results-worse-than-2013-Umalusi-20141230 (reference)
7. Breiteig T, Grevholm B. The transition from arithmetic to algebra: To reason, explain, argue, generalize and justify. In: Novotná J, Morová H, Krátká M, Stehlíková NA, editors. Proceedings of the 30th Conference for the International Group for the Psychology of Mathematics Education. Prague: Charles University Prague; 2006. p. 2-225–2-232.
8. Kaput J. A research base supporting long term algebra reform. In: Owens DT, Reed K, Millsops GM, editors. Proceedings of the 17th Annual Meeting of the North America Chapter of PME. Volume 2. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education; 1995. p. 71–108.
9. Moseley B, Brenner ME. A comparison of curricular effects on the integration of arithmetic and algebraic schemata in pre-algebra students. Instr Sci. 2009;37:1–20. https://doi.org/10.1007/s11251-008-9057-6
10. Carroll C. Evaluation of the University of Limerick Mathematics Learning Centre [BSc report]. Limerick: University of Limerick; 2011.
11. Sheridan B. How much do our incoming first year students know?: diagnostic testing in Mathematics at third level. Ir J Acad Pract. 2013;2(1), Art. #3, 19 pages. Available from: https://arrow.dit.ie/cgi/viewcontent.cgi?referer=https://www.google.co.za/&httpsredir=1&article=1010&context=ijap
12. Maharaj A, Wagh V. An outline of possible pre-course diagnostics for differential calculus. S Afr J Sci. 2014:110(7/8), Art. #2013-0244, 7 pages. http://dx.doi.org/10.1590/sajs.2014/20130244
13. Makuuchi M, Bahlmann J, Friederici AD. An approach to separating the levels of hierarchical structure building in language and mathematics. Phil Trans R Soc B. 2012;367:2033–2045. https://doi.org/10.1098/rstb.2012.0095
14. Groebner DF, Shannon PW, Fry PC, Smith KD. Business statistics: A decision making approach. 6th ed. Upper Saddle River, NJ: Pearson Education Inc.; 2005.
15. University of KwaZulu-Natal (UKZN). College of Agriculture, Engineering and Science handbook. Pietermaritzburg: College of Agriculture, Engineering and Science, UKZN; 2015.
16. Maharaj A, Brijlall D, Narain O. Designing website-based tasks to improve proficiency in mathematics: A case of basic algebra. Int J Educ Sci. 2015;8(2):369–386. https://doi.org/10.1080/09751122.2015.11890259

The Grootfontein aquifer: Governance of a hydro-social system at Nash equilibrium

AUTHORS:
Jude E. Cobbing[1]
Maarten de Wit[1]

AFFILIATION:
[1]AEON/ESSRI, Nelson Mandela University, Port Elizabeth, South Africa

CORRESPONDENCE TO:
Jude Cobbing

EMAIL:
jcobbing@gmail.com

DATES:
Received: 05 Aug. 2017
Revised: 26 Sep. 2017
Accepted: 04 Jan. 2018
Published: 30 May 2018

KEYWORDS:
groundwater; management; institutions; transdisciplinarity; resilience

HOW TO CITE:
Cobbing JE, De Wit M. The Grootfontein aquifer: Governance of a hydro-social system at Nash equilibrium. S Afr J Sci. 2018;114(5/6), Art. #2017-0230, 7 pages. http://dx.doi.org/10.17159/sajs.2018/20170230

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Water Research Commission (South Africa)

The Grootfontein groundwater aquifer is important to the water supply of the town Mahikeng in the North West Province of South Africa and to commercial agriculture in the Province, but the water table has fallen by up to 28 m as a consequence of over-abstraction since the 1980s. Institutional and hydrogeological issues impact the aquifer in complex ways, described here as a hydro-social system. Whilst the hydrogeology is well understood and South African laws provide for sustainable groundwater governance, poor stakeholder collaboration and other institutional problems mean that the over-abstraction is likely to persist – an example of an undesirable Nash equilibrium. The Grootfontein aquifer case shows that groundwater underpins wider social-ecological-economic systems, and that more holistic management – taking the institutional context into account – is needed to underpin economic growth, employment and other public outcomes.

Significance:

- The cost of better natural resource stewardship, including groundwater, is likely to be considerably less than the losses that occur when it is absent.
- If local groundwater was better managed, it could make water supplies in Mahikeng cheaper and more reliable, which would in turn support local economic growth and employment.

Introduction

Groundwater's global social, economic and environmental importance contrasts with its low profile amongst policymakers and the general public. It is the primary domestic water source for about half of the global population[1], and for about three quarters of Africa's population[2]. More than half of all South Africans depend on groundwater for their domestic water supply[3], and the total renewable volume of groundwater in South Africa is similar to the surface water assured yield[4,5]. Groundwater cannot deliver large volumes at a single location like a dam, but its dispersed nature is an advantage in supplying rural areas and scattered small settlements with potable water. Whilst South Africa's surface water resources are nearly fully allocated, only around half of the national renewable groundwater resource is used.[5,6]

Unfortunately, groundwater supply systems in South Africa are often poorly maintained[7], and groundwater management is inadequate[8,9]. Municipalities tend to prefer surface water over groundwater.[7] Here we examine a well-studied and prolific South African aquifer – the Grootfontein dolomite aquifer in the North West Province – to investigate poor groundwater management and to discern wider lessons for social-ecological-economic stewardship in South Africa.

The North West dolomites, which stretch from Gauteng Province to the border with Botswana, are amongst South Africa's most important and prolific groundwater aquifers.[10,11] They cover an area of about 5000 km^2, hold a renewable water resource of similar magnitude to the Gariep Dam, and receive natural recharge of about 300 Mm^3/a.[12,13] This groundwater resource supports extensive agricultural irrigation and the water supplies of towns such as Itsoseng, Lichtenburg, Mahikeng, Ottoshoop, Ventersdorp and Zeerust, as well as hundreds of dispersed settlements and homesteads. The North West dolomites are sub-divided into a patchwork of semi-autonomous units or 'compartments' by geological features such as igneous dykes.[12,14] Under natural conditions, each compartment discharges groundwater via springs and wetlands, and is recharged by rainfall. One of these compartments is the Grootfontein compartment or aquifer, for which inadequate groundwater management has contributed to falling water levels.

The Grootfontein aquifer, which is 30 km southeast of the town of Mahikeng, was Mahikeng's main water source until the early 1980s. Today it supplies only about 20% of the town's water (about 10 ML/day) because its yield has declined and Mahikeng's demand has grown.[15] The rest of Mahikeng's water comes from a large spring, the Molopo Eye, which drains a different dolomite compartment to the north of the Grootfontein aquifer (about 20 ML/day), and the Setumo Dam on the ephemeral Molopo River, downstream of Mahikeng (another 20 ML/day); see Figure 1.

Problem statement

The Grootfontein aquifer once discharged naturally at its northern (down-gradient) boundary, mainly from a large spring (the 'Grootfontein', or great spring). The drilling of boreholes around the spring to increase Mahikeng's supply in the 1970s, combined with large increases in irrigation abstractions elsewhere in the compartment, finally dried up the spring in 1981. As water levels in the aquifer dropped further, some of the boreholes around the spring failed. Today the groundwater level in the Grootfontein aquifer near the former spring is more than 28 m below ground level.[15] The three remaining viable public water supply boreholes at the former spring now yield less than half of the original combined wellfield potential (i.e. only about 10 ML/day). Irrigating farmers, who abstract the lion's share of the groundwater, are concerned about falling water levels, higher pumping costs and growing uncertainty.

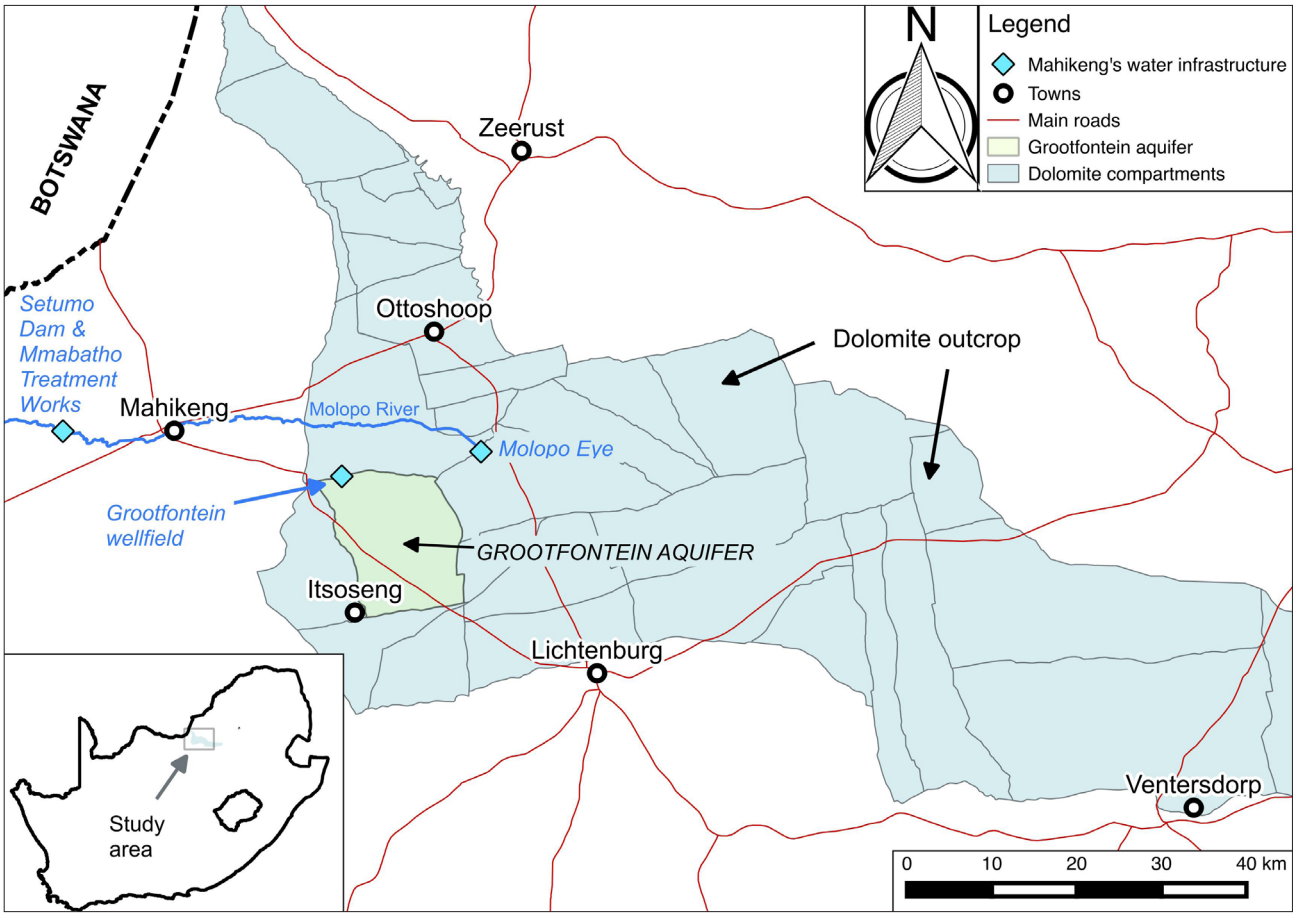


Figure 1: The Grootfontein aquifer and the dolomite outcrop near Mahikeng (boundaries after Holland and Wiegmans[14]).

Partly in response to the uncertain supply from Grootfontein, Mahikeng city officials have invested more than ZAR30 million in upgrading the water treatment plant at the Setumo Dam, from a supply of 10 ML/day to today's 20 ML/day.[16] The Setumo Dam receives a large proportion of its inflow from discharges from Mahikeng's two wastewater treatment plants, and from leaks in the town's reticulation system. The dam water quality is consequently poor, requiring sophisticated treatment to attain South African drinking water quality standards.[16] In contrast, Mahikeng's groundwater sources are of generally good quality, needing only prophylactic chlorination before entering supply. As the dam depends partly on return flows from the town, and the Molopo Eye spring is vulnerable to small changes in water level, the Grootfontein aquifer is more important than it might appear.

Hydrogeologically, the Grootfontein aquifer is amongst the best-studied aquifers in South Africa[17], with technical research dating back to the 1960s or before[18,19]. At least 15 of the South African Department of Water and Sanitation (DWS)'s technical geohydrology reports concentrate on Grootfontein.[17] Some hydrogeological uncertainty remains, but in general Grootfontein is technically well understood.[15]

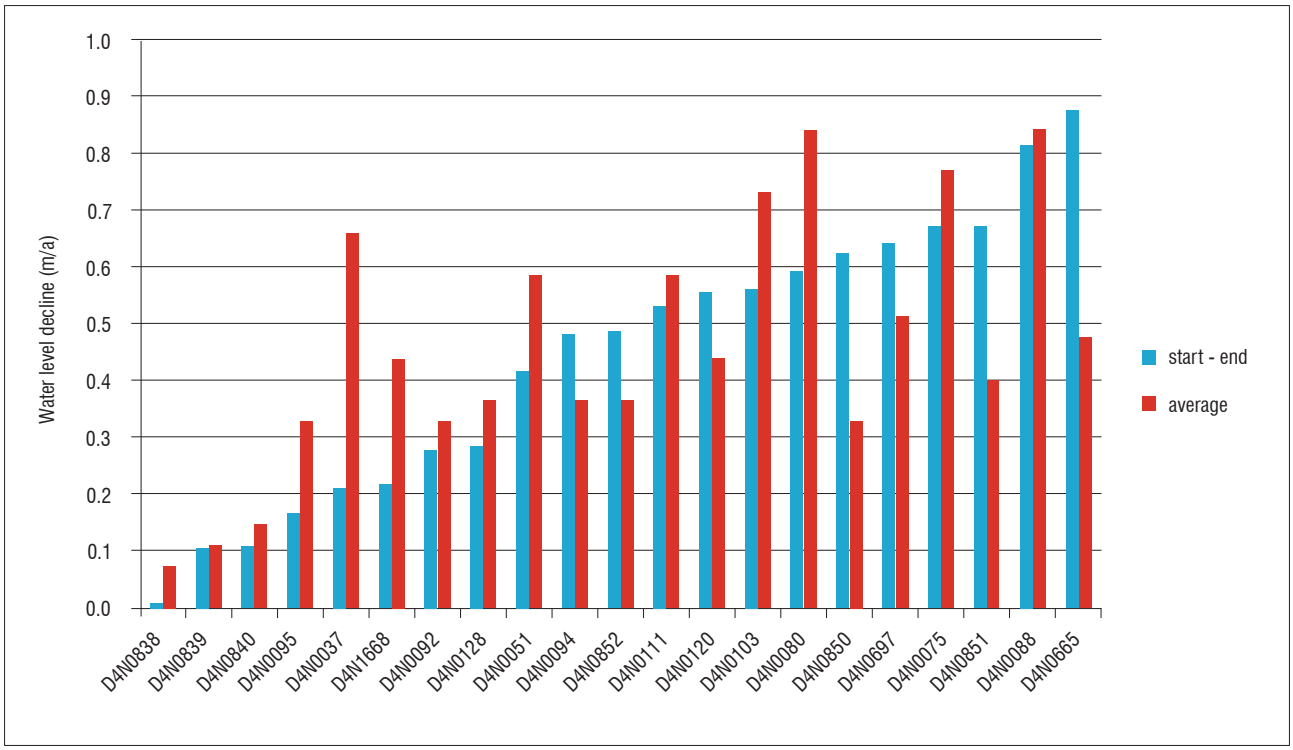
Groundwater levels at Grootfontein have been falling since the early 1980s, despite the relatively thorough hydrogeological understanding of the aquifer,[15,17,20] its proximity and importance to a provincial capital, and the legal mechanisms intended to prevent overuse. Figure 2 summarises the records of 21 DWS groundwater-level monitoring stations within the Grootfontein aquifer. Blue columns show the difference between the mean of the first year's readings and the mean of the last year's readings for each record (about 0.44 m/a on average), and red columns show average decline based on a straight-line fit through each record (about 0.46 m/a on average).[15,17] Station record lengths vary between 16 and 43 years. Higher average declines occur near the Grootfontein wellfield.[17] This overall decline has caused the failure of some boreholes abstracting water for Mahikeng, has increased costs and uncertainty for irrigating farmers, and has deprived Mahikeng of the reserve of groundwater that a higher saturated aquifer thickness would allow.

Sound technical knowledge of the aquifer, and legal provisions designed to prevent over-exploitation, plainly do not ensure sustainable use of Grootfontein groundwater. The question therefore arises: What is necessary for the successful management of the Grootfontein aquifer?

Methodology

Research carried out between 2013 and 2015 aimed to understand the reasons for poor groundwater governance at Grootfontein. The existing extensive hydrogeological knowledge of Grootfontein implies that social and institutional dynamics underpin poor governance, rather than lack of technical knowledge. However, we also investigated the current hydrogeological understanding of the aquifer, because this understanding delineates physical limits to management interventions. A mixed-methods research approach was therefore used[21], which combined social science and physical science techniques, in keeping with the principles of earth stewardship science[22].

This research approach combined conventional hydrogeological methods (literature review, field sampling, and analysis of hydraulic parameters leading to a conceptual model and water balance), with 63 field interviews and participant observation.[17] Participant observation means taking part in meetings as a recognised participant, for example a meeting called by DWS to discuss the aquifer, or a municipal water infrastructure tender briefing. This participation helped illuminate the day-to-day working environments of the main organisations involved in groundwater governance at Grootfontein, and the formal and informal interactions between these organisations. A combination of quantitative evidence (e.g. water levels, pumping volumes, chemical analyses) and qualitative evidence (e.g. interview responses, or the absence of a groundwater protection zone around a borehole) was used to corroborate conclusions in a process of triangulation.[21]



Data source: South African Department of Water and Sanitation

Figure 2: Rates of water level decline for the Grootfontein aquifer.

It became clear that both the social/institutional *and* the hydrogeological/technical aspects of the aquifer's functioning were important to any attempt to address the complexity of this resource. Indeed, the situation at Grootfontein can be characterised as a 'hydro-social system', meaning that these aspects influence each other in relatively complex and often counter- and non-intuitive ways.

Whilst the social and institutional aspects appear to control the eventual outcomes regarding water use and water levels, the hydrogeology defines what is physically possible. Aspects of the hydrogeology also heavily influence the institutional features of the system. For example, the heterogeneity of hydraulic properties as a function of complex weathering and karst formation reinforces a common notion of the dolomite groundwater resource as mysterious and unreliable. We argue that knowledge of both the institutional and the hydrogeological aspects is necessary in order to understand the hydro-social system. No hierarchy is implied.

Results

Overview of the governance of the Grootfontein aquifer

Sweeping changes to South African water law since 1997 mean that ownership of water is now vested in the state via the Department of Water and Sanitation (DWS). The law specifies minimum volumes of water for environmental functioning and for basic human needs, which take precedent over other uses.[23] All other water uses must be licensed. In practice, verification of licenced quantities (e.g. of irrigation boreholes) is rare, and penalties for over-abstraction are uncommon. The principle of subsidiarity or decentralisation is prescribed[24], but most of the envisaged basin-level and local-level organisations have not emerged since the new legislation, which is now being revised[25].

In particular, the *National Water Act*[23] envisages Water User Associations (WUAs) as cooperative local associations of individual water users that manage water along integrated water resource management lines. WUAs were intended to replace the apartheid-era irrigation boards, emphasising equitable access to water and improving social outcomes. Constitution of the WUAs was very slow, with DWS concerned that proposed WUAs merely replicated former structures. No WUAs were ever approved for the aquifers of the North West dolomites, and today DWS no longer supports WUAs and is disbanding those that do exist elsewhere in the country.[25]

Major stakeholders at Grootfontein with a direct influence on the groundwater abstractions include DWS, the irrigating farmers, the local and district municipalities, and the regional water board. Technical and policy advisors and consultants, and other government departments, also have a role. As described, semi-structured interviews and participant observation methods were used to gain insight into these stakeholders,[17] which are briefly summarised here:

1. The two municipalities (Mahikeng Local Municipality and Ngaka Modiri Molema District Municipality, or NMMDM) are challenged by issues including internal organisation, budgeting, mandate and skills, and consequently struggle to fulfil their mandates. In 2014, NMMDM was placed under provincial administration.[26]
2. The regional water board, Sedibeng Water Board, has considerable technical resources and has received unqualified audits since 2002.[27] It took over the underperforming Botshelo Water Board in early 2015. Sedibeng Water Board's responsibilities in Mahikeng include the two water treatment plants, the Setumo Dam, both wastewater treatment plants, and a portfolio of water reticulation, storage and metering assets. It also has responsibility for large rural areas with many small groundwater sources. It is focusing on the upgrade to the Mmabatho Water Treatment Works; management of the primary groundwater sources (the Grootfontein wellfield and the Molopo Eye) is carried out by DWS.
3. DWS, as the legal custodian of all water in South Africa, has ultimate responsibility for water resources such as Grootfontein, where it operates and manages the wellfield. It has an office in Mahikeng, and a small satellite office at the Grootfontein wellfield. In 2015, DWS was working on the proposed nine Catchment Management Agencies, with final responsibilities for licensing and control of local groundwater resources still to be resolved. DWS opposes WUAs because they are thought to replicate undesirable past modes of governance, but this policy has unintentionally contributed to the lack

of effective local-level groundwater management organisations.[17] The emerging Catchment Management Agencies will eventually assume the WUA responsibilities, but in the meantime these functions are partly vested in Catchment Management Forums and Stakeholder Operating Forums hosted and organised by DWS. In general, these interim forums do not attract wide stakeholder support in Mahikeng, and are convened by relatively junior DWS personnel.[17] Issues of internal organisation and mandate also challenge DWS,[28] which in turn contributes to poor availability of internal hydrogeological data, inadequate protection measures at the Grootfontein wellfield, and poor verification and validation of groundwater abstraction licences.

4. Irrigating farmers at Grootfontein abstract most of the groundwater, mainly to feed centre-pivot irrigation systems.[17] No effective forum exists to manage these groundwater abstractions, and irrigation abstractions are limited primarily by infrastructure characteristics, electricity prices and irrigation requirements. Farmers are not as homogeneous a group as is sometimes assumed, and farm incomes depend on groundwater irrigation to varying degrees. In some cases, irrigation could be substantially reduced with proportionately little impact on farm income, such as by changing crop types or by allowing fields to lie fallow in the dry winter months.
5. Other stakeholder groups at Grootfontein include technical consultants, international policy experts and other government departments (e.g. National Treasury) which, although they have little direct stake in Grootfontein groundwater, have significant long-term influence on groundwater policy and operations. On the other hand, numerous local households and small businesses (e.g. at Itsoseng) depend on Grootfontein groundwater; these users abstract small amounts that are critical to livelihoods, but have little influence on abstraction policy.

Analysis of governance at Grootfontein

Irrigating farmers at Grootfontein are reluctant to unilaterally reduce abstractions, without a similar gesture from DWS or from NMMDM as the water services provider. DWS and NMMDM consider some farmers to be in breach of abstraction licence agreements, but are reluctant to engage in legal or other actions to address this breach – a situation that is complicated by ongoing organisational changes at DWS, technical difficulties in assessing abstraction volumes, and poor relations between DWS and NMMDM.[17] No effective local forum exists in which 'win-win' situations are debated – a situation exacerbated by the demise of WUAs in the North West dolomites and the inadequate interim replacements.

Misperceptions about the aquifer potential are worsened by a lack of accessible abstraction and other hydrogeological data, although these data do exist. Institutional bias towards surface water inclines many stakeholders (including DWS) towards upgrading the Mmabatho Water Treatment Works at the Setumo Dam[16], and even a pipeline from another catchment[17]. These surface water solutions also allow city officials to avoid tackling groundwater governance problems.

An analysis of the various stakeholders at Grootfontein was carried out, using the six institutional characteristics or 'appropriator attributes' suggested by Ostrom[29] as a framework[17]. These six appropriator attributes are[29]:

1. **Salience**: Appropriators depend on the resource system for a major portion of their livelihood or the achievement of important social or religious values.
2. **Common understanding**: Appropriators have a shared image of how the resource system operates and how their actions affect each other and the resource system.
3. **Low discount rate**: Appropriators use a sufficiently low discount rate in relation to future benefits to be achieved from the resource.
4. **Trust and reciprocity**: Appropriators trust one another to keep promises and relate to one another with reciprocity.
5. **Autonomy**: Appropriators are able to determine access and harvesting rules without external authorities countermanding them.
6. **Prior organisational experience and local leadership**: Appropriators have learned at least minimal skills of organisation and leadership through participation in other local associations or through learning from approaches that neighbouring groups have taken.

If present in a group of common pool resource appropriators, these appropriator attributes '…enhance the likelihood of appropriators organising themselves to try to avoid the social losses associated with open access or rules that are not yet working well'[29]. It was found that the hydro-social system at Grootfontein does not possess any of these appropriator attributes, making it less likely that stakeholders would naturally organise themselves or collaborate to manage the aquifer sustainably.[17] These appropriator attributes depend not only on the current characteristics of the appropriators (or stakeholders), but also on the history of interactions between them and various other factors. According to Ostrom, attributes are '…affected by the larger regime in which a resource and its appropriators are embedded'[29].

The current situation can be described as a sub-optimal equilibrium or stalemate, in which major abstracting stakeholders have no incentive to reduce abstractions.[17] Instead, stakeholders maximise abstractions whilst the resource lasts – a classic 'tragedy of the commons'[30]. Sub-optimal equilibria of this sort are sometimes known as Nash equilibria – a term originating in mathematical game theory and referring to situations in which 'players' perceive no advantage to changing their behaviour, despite a collective long-term sub-optimal outcome.[31,32] As Nasar[31] puts it:

> *Nash equilibria – defined as each player's following his best strategy assuming that the other players will follow their best strategy – aren't necessarily the best solution from the vantage point of the group of players.*

In order to break a Nash equilibrium and precipitate change, one 'player' or stakeholder needs to 'make a move' that may be disadvantageous in the short term.

An effective local forum needs to intervene, and be backed by good data availability and with powers of coercion and (if necessary) enforcement. WUAs were originally conceived as such, but were never constituted. Of the major stakeholders, DWS has the best chance of convening, forming or otherwise presiding over such a forum.[33] DWS is legally mandated to assume such a role, whereas other stakeholders are unlikely to intervene unilaterally. Irrigating farmers, the major users of Grootfontein groundwater, see eventual failure of the resource as possible, but have neither the internal organisational attributes nor external incentives to change their behaviour.

The slow change from DWS regional offices to Catchment Management Agencies, the perception that the issues are being addressed by the Stakeholder Operating and Catchment Management Forums, internal organisational difficulties (e.g. in accessing data) and a lack of support from senior management mean that DWS is unlikely to make significant changes to the situation in the near future. The apportionment of blame for the situation, linked to the false promise of 'hard-state' authoritarian legal remedies[34], feeds a zero-sum mentality and detracts from the complex and incremental task at hand[17]. Thus over-abstraction at Grootfontein is likely to continue.

The analysis described above has implications for hydrogeological research in South Africa, and also for the wider social, environmental and economic context. These implications are discussed in the next section.

Implications of the findings

Implications for hydrogeological research

The Grootfontein aquifer is well understood technically, and could provide a larger and more reliable supply of good quality water to Mahikeng.[15] Analysis of the complex web of organisations and stakeholders outlined

above suggests that the crux of the over-abstraction problem lies not in poor technical knowledge, but is in fact a more complex social and institutional issue. The problem is nevertheless often wrongly framed, explicitly or implicitly, as related to a lack of data or to the vagaries of a fundamentally capricious and unreliable aquifer. Calls for better understanding of the recharge, or more explicit delineation of the aquifer boundaries, inadvertently reinforce this misperception and detract from the central issue, which is that governance arrangements are inadequate.

The interdependence of the hydrogeology and the institutional framework for governance means that groundwater governance at Grootfontein (and elsewhere) can be described as a political economy issue. Political economy analysis examines institutional interactions and functioning in the context of how power is apportioned and finds traction.[35] The technical (or hydrogeological) characteristics are only one component of the political economy of a groundwater supply system, but political economy issues are often ignored by technical specialists.

However, classic political economy analysis can place too little emphasis on the technical hydrogeology. At worst it tends to assume that water governance can be understood by examining only the institutional, economic and material circumstances of the participants, with the influence and feedback of the hydrological sphere relegated to the margins. It is convenient to see all water resources as essentially similar. The institutional or social sphere must in fact evolve alongside the physical characteristics of the hydrological system, or risk collapse when underlying hydrological assumptions prove false.

Hydrogeological characteristics provide a boundary of hydraulic possibility, but other factors control who abstracts how much groundwater, and when. This implies that any final 'safe yield' figure at Grootfontein (as elsewhere) will essentially be politically mediated – the ideal hydraulic figure may differ from the best figure in the wider institutional sense, particularly when multi-year aquifer changes are under consideration, and refinements to hydraulic figures are in any case consequent on feedback from 'adaptive management' arrangements.[9,36]

Most hydrogeological research at Grootfontein has been state funded, which implies a fundamental public interest. If hydrogeological research cannot be explicitly linked to public interest outcomes, then justifying and funding such research will become increasingly difficult. This in turn implies that hydrogeologists (and other technical specialists) should understand and even explicitly account for the political economy or institutional context of their work. Ignoring this context may damage the link between scientific research and public policy, or worse, erode the social licence to operate as publicly mandated researchers.

Implications for the environment and the economy

As Grootfontein illustrates, important groundwater supplies in South Africa have inadequate governance systems that enable over-abstraction. Inadequate groundwater governance in turn harms economic growth, food security, social stability, land reform, transformation and other sectors.

Parts of Mahikeng are supplied with diluted wastewater, which is treated to obtain drinking water standards at the multimillion rand Mmabatho Water Treatment Works. At the same time, groundwater of excellent quality is being used to irrigate crops, including maize destined for cattle feed. Apart from the higher cost of treating wastewater, Mahikeng is dependent on a high-tech treatment plant that has failed in the past[16] and may fail again. This situation also lowers the resilience of Mahikeng's water supply system as a whole by increasing its vulnerability to unexpected shocks or outages. Mmabatho Water Treatment Works is technically sophisticated and well engineered, but all such systems depend on long institutional chains involving funding, staffing, consumables supply, maintenance and financing. Utilising the superior water quality and storage of the Grootfontein and other dolomite compartments would increase overall system resilience by incorporating passive systems requiring lower energy and capital inputs, such as back-up storage underground. The resilience of the town's water supply system in turn impacts on the perceptions of residents, investors and businesses, and on the wider resilience of its economic and social functioning.

Environmental impacts of groundwater over-abstraction in the North West dolomites include the destruction of several large springs and associated wetlands and ecosystems (e.g. the Grootfontein Eye, the Lichtenburg Eye and the Polfontein Eye at Itsoseng).

The Grootfontein hydro-social system is therefore a component of wider regional social-ecological-economic systems, each with a complex range of intermingled institutional characteristics as well as bounding technical or physical limits. Such systems are difficult to depict and characterise, and are inherently transdisciplinary. They are also dynamic in nature, and evolve in response to diverse forces. The discourse of 'resilience' is useful when considering this bigger picture.

Implications for resilience

Arising out of ecological studies and the sustainability discourse, resilience refers to the ability of a dynamic system to absorb shocks, as well as to the 'capacity for renewal, re-organisation and development'[37]. Human societies are ultimately dependent on their ecological context[38], and the profit-driven externalising of environmental damage may be ultimately self-limiting on aggregate[39]. A resilient social-ecological-economic system is one that adapts to shocks (e.g. droughts, pollution, commodity price collapse) and persists, whilst minimising adverse consequences (e.g. a rise in unemployment, or long-term ecological damage).

The economy is also important to resilience – economic dynamism is a pre-requisite for the political and social stability required to overcome complex long-term environmental problems. In some cases, resilience (and linked fields such as sustainability science) can put the cart before the horse, thereby reflecting a myth more common in the developed world that economic progress *requires* parallel environmental sustainability. Historically, economic development is more closely correlated with environmental devastation, at least in the short and medium term. Today, many economists believe that economic development can and should be delinked from unsustainable environmental exploitation, and there is an ethical case for this belief. Environmental devastation is also possible without catalysing economic development – an outcome unfortunately common across Africa.

Social-ecological-economic systems that are poorly understood, or are simplistically outlined based on convenient or politically expedient variables, are more vulnerable to unexpected 'cascades of failure' that can occur in complex systems in which linkages and feedback mechanisms are ill defined.[40,41] A limited understanding of such systems reduces resilience as effectively as concrete factors such as drought or lack of finance, as it prevents optimal resource allocation and increases risk.

The problem is not unique to Grootfontein. We briefly describe two further examples that illustrate the linkages between poor groundwater governance in the North West dolomites and the wider social-ecological-economic context:

1. Over-abstraction of groundwater from the dolomite aquifers in and around the town of Lichtenburg has been scientifically described since the 1960s,[42] but never resolved. It has contributed to domestic water shortages, burning of peat deposits, concerns over the viability of dairying, irrigated agriculture, cement production and other economic activities, and the drying up of the town spring (the Lichtenburg Eye) and associated public amenity. It also contributes to a corrosive sense that the designated authorities fail to act in the public interest more generally. In the short term, these things incentivise further over-abstraction, also seen at Grootfontein. Long term, they fuel lack of investment and unemployment.
2. Poor management of groundwater levels in the Steenkoppies dolomite compartment near Krugersdorp threatens the viability of valuable irrigated agriculture as well as businesses and ecosystems downstream that depend on Magalies River water.[43,44] It has led to *de facto* 'management by court order' in which potential partners in collaborative management shun each other, or threaten each other in court. The situation militates against the kind of economic dynamism, social cohesion and catalysing of opportunity necessary for the step-increase in the numbers of dignified, decently paid jobs

envisaged by the National Development Plan.[45] Recognising key linkages between groundwater and other sectors is a first step to improving governance, and ultimately promoting a more resilient social-ecological-economic system at Steenkoppies.

Policy discussion

A key factor in North West dolomites groundwater governance problems is the lack of true collaboration between local stakeholders and central authority. DWS's decision to withhold official recognition of WUAs in the North West dolomites unintentionally contributed to the breakdown of formal local groundwater governance mechanisms, the *de facto* continuation of past modes of water use and governance, and a general cynicism and short-termism. This decision is justified on the grounds that the incipient WUAs were socially regressive, but the lack of an authoritative, speedy and effective replacement has inadvertently worsened the problem. New policies must be put into practice. As Aarnoudse et al.[46] put it:

> *New policies are not institutions as such. They first need to relate to existing patterns and structures and add to, synergise with or replace them in order to achieve institutional change.*

Decentralised modes of governance imply theoretically lower transaction costs, better consultation and faster decisions, but a central authority is also needed to break logjams, and engage with, endorse and hold accountable local governance decisions in line with the democratic mandate.[47] A central authority has a wider perspective of regional social-ecological-economic systems, and can if necessary make decisions that increase overall resilience even when this might oppose parochial or short-term interests (e.g. in breaking local Nash equilibria). The *National Water Act* explicitly recognises this function of central government.[23,24] DWS has not always played this role in the North West dolomites[33]; even the verification and validation of large volume groundwater abstraction licences – a first step in effective governance – has faltered.

Collaboration between government departments (e.g. DWS; the Department of Rural Development and Land Reform; the Department of Agriculture, Forestry and Fisheries; the Department of Economic Development; and the National Treasury) and across spheres of government (e.g. between municipalities, provincial government and national government) is inadequate when it comes to groundwater governance, despite this resource underpinning many explicit goals of the various sectors. Such collaboration is essential if optimal social outcomes are to be achieved. The general problem of poor cross-sectoral coordination in South Africa has been recognised in the National Development Plan[45], but is particularly acute in the field of groundwater governance because groundwater is hidden from view, both physically and institutionally.[17]

The costs of less-resilient or more vulnerable social-ecological-economic systems, whilst difficult to quantify, are much higher than the costs incurred by effective management. The risks of water supply failure, with unpredictable implications, are also disproportionately borne by the poor and the vulnerable. Wealthier individuals and businesses can insulate themselves against water supply uncertainty (e.g. by installing on-site storage or household reverse-osmosis treatment systems), but in the long run society prospers or fails together.

In the absence of effective policy and action, there is an implicit endorsement of the potentially regressive and short-term modes of environmental governance that arise ad hoc, and that often owe much to past laws or past forms of social and demographic interaction. A 'business as usual' approach, based either on the absence of policy or on the division of the social-ecological-economic sphere into silos, is unlikely to bring about the kind of step change in environmental governance (and socio-economic development) that is required to break the Nash equilibrium.

Conclusions

Groundwater over-exploitation at Grootfontein is neither natural nor inevitable, but is the collective result of stakeholder actions and choices over many years. These actions are based not only on laws and formal water governance policies, but rely also on hierarchies of power, historical modes of thought and perception, short-term socio-political incentives, interference from other sectors (e.g. agriculture and land reform), and other institutional factors. These factors are less tangible and more controversial than the physical hydrogeology or water resource.

The Grootfontein hydro-social system responds to natural changes (e.g. droughts and recharge events) and to social or institutional pressures (e.g. over-pumping, availability of infrastructure funding, political priorities). Better management requires an understanding of both. The Grootfontein hydro-social system is in turn a component in wider social-ecological-economic systems, whose resilience is important for national development and public outcomes.

Acknowledgements

We thank the Water Research Commission for funding part of this research, interviewees in Mahikeng and elsewhere, and hydrogeologists at the Department of Water and Sanitation.

Authors' contributions

J.E.C. was responsible for the conceptualisation, methodology, data collection, data analysis, sample analysis, data curation, writing the initial draft, writing revisions, project management and funding acquisition. M.d.W. was responsible for the conceptualisation, critically reviewing the writing, student supervision and project leadership.

References

1. Alley WM, Alley R. High and dry. Meeting the challenges of the world's growing dependence on groundwater. New Haven, CT: Yale University Press; 2017.
2. United Nations Economic Commission for Africa (ECA). The Africa water vision for 2025: Equitable and sustainable use of water for socioeconomic development. Addis Ababa: ECA; 2004.
3. Braune E, Xu Y. A South African perspective on the protection of groundwater resources. In: Xu Y, Usher B, editors. Groundwater pollution in Africa. London: CRC Press; 2006. https://doi.org/10.1201/9780203963548.ch29
4. South African Department of Water Affairs (DWA). National water resource strategy. Pretoria: DWA; 2004.
5. South African Department of Water Affairs (DWA). Groundwater strategy 2010. Pretoria: DWA; 2010.
6. Woodford A, Rosewarne P, Girman J. How much groundwater does South Africa have? Pretoria: SRK Consulting; 2006. Unpublished report.
7. Cobbing JE, Eales K, Gibson J, Lenkoe K, Cobbing BL. Operation and maintenance (O&M) and the perceived unreliability of domestic groundwater supplies in South Africa. S Afr J Geol. 2015;118(1):17–32. https://doi.org/10.2113/gssajg.118.1.17
8. Pietersen K, Beekman HE, Holland M. South African groundwater governance case study. Report prepared for the World Bank in partnership with the South African Department of Water Affairs and the Water Research Commission. WRC report no. KV 273/11. Pretoria: Water Research Commission; 2011.
9. Knuppe K. The challenges facing sustainable and adaptive groundwater management in South Africa. Water SA. 2011;37(1):67–79. https://doi.org/10.4314/wsa.v37i1.64110
10. Barnard HC. An explanation of the 1:500 000 General Hydrogeological Map Johannesburg 2526. Pretoria: Department of Water Affairs and Forestry; 2000.
11. Vegter JR. Groundwater development in South Africa and an introduction to the hydrogeology of groundwater regions. WRC report no. TT134/00. Pretoria: Water Research Commission; 2001.
12. Meyer R. Hydrogeology of Groundwater Region 10: The Karst Belt. Water Research Commission report no. TT 553/12. Pretoria: Water Research Commission; 2012.
13. Stephens A, Bredenkamp DB. Institutional arrangements for groundwater management in dolomitic terrains: Situation analysis. WRC report no. KV140/02. Pretoria: Water Research Commission; 2002.

14. Holland M, Wiegmans F. Geohydrology guideline development: Activity 18&19 desktop development of a dolomite hydrogeological compartment map and explanation booklet (report). Report prepared for the Department of Water Affairs by Water Geosciences Consulting as part of DWA project number 14/14/5/2: Implementation of dolomite guideline. Pretoria: Department of Water Affairs; 2009.

15. Cobbing JE. An updated water balance for the Grootfontein aquifer near Mahikeng. Water SA. 2018;44(1):54–64.

16. South African Department of Water and Sanitation (DWS). Third quarter performance evaluation report for Greater Mafikeng Bulk Water Supply Scheme. Period 1 October to 31 December 2014. Pretoria: Chief Directorate: Bulk Infrastructure Programme, Department of Water and Sanitation; 2014.

17. Cobbing JE. The Grootfontein aquifer as a hydro-social system [PhD thesis]. Port Elizabeth: Nelson Mandela Metropolitan University; 2017.

18. Bredenkamp DB. Verslag van hidrologiese opname in die Bo-Molopo Ondergrondse Waterbeheergebied [Report of hydrological surveys in the Upper Molopo Subterranean Water Control Area]. Report no. GH 1283. Pretoria: Department of Water and Sanitation; 1964. Afrikaans.

19. Temperley BN. Groundwater conditions on Spring Valley Farm, Rooigrond near Mafikeng. Report no. GH 1273. Pretoria: Department of Water and Sanitation, Pretoria; 1965.

20. Van Tonder GH, Janse van Rensburg H, Botha JF, Bredenkamp DB. Die modellering van grondwatervlakke in die Grootfonteinkompartement in Wes-Transvaal [The modelling of groundwater levels in the Grootfontein compartment in the Western Transvaal]. Water SA. 1986;12(3):151–160. Afrikaans.

21. Yin RK. Case study research. Design and methods. 4th ed. Los Angeles, CA: Sage Publications; 2009.

22. De Wit MJ, Booth P. Iphakade is Earth Stewardship Science. S Afr J Geol. 2016;119(1):3–14. https://doi.org/10.2113/gssajg.119.1.3

23. RSA. Act No. 36 of 1998: National Water Act. Cape Town: Republic of South Africa Government Gazette 398 (19182); 1998.

24. Lazarus P. Towards a regulatory framework for the management of groundwater in South Africa. Water Research Commission report no. 789/1/98. Pretoria: Water Research Commission; 1998.

25. South African Department of Water and Sanitation (DWS). National Water Policy Review (NWPR). Approved water policy positions, 31 January 2014. Pretoria: DWS; 2014.

26. Ngaka Modiri Molema District Municipality (NMMDM). Ngaka Modiri Molema District Municipality Annual Report for the 2014/2015 Financial Year. Mahikeng: NMMDM; 2015.

27. Annual report of Sedibeng Water Board for the 2014/15 financial year. Mahikeng: Sedibeng Water Board; 2015.

28. Seward P. Rethinking groundwater governance in South Africa [PhD thesis]. Cape Town: University of the Western Cape; 2015.

29. Ostrom E. Understanding institutional diversity. Princeton, NJ: Princeton University Press; 2005.

30. Hardin G. The tragedy of the commons. Science. 1968;162(3859):1243–1248. https://doi.org/10.1126/science.162.3859.1243

31. Nasar S. A Beautiful Mind. The life of mathematical genius and Nobel laureate John Nash. New York: Simon and Schuster; 1998.

32. Nash JF. Equilibrium points in n-person games. Proc Natl Acad Sci USA. 1950;36(1):48–49. https://doi.org/10.1073/pnas.36.1.48

33. Riemann K, Rust J, Hoosain M, Jack S. Water governance framework and action plan for conjunctive use. Draft final deliverable of WRC project K5/2332: Water governance of groundwater and surface water resources in South Africa. Pretoria: Water Research Commission; 2016.

34. Mukherji A, Shah T. Groundwater socio-ecology and governance: A review of institutions and policies in selected countries. Hydrogeol J. 2005;13:328–345. https://doi.org/10.1007/s10040-005-0434-9

35. Manghee S, Poole A. Approaches to conducting political economy analysis in the urban water sector. World Bank Water Paper 74742. Washington DC: World Bank Group; 2012.

36. Seward P, Xu Y, Brendonck L. 2006 Sustainable groundwater use, the capture principle, and adaptive management. Water SA. 2006;4:473–482.

37. Folke C. Resilience: The emergence of a perspective for social-ecological systems analyses. Glob Environ Change. 2006;16:253–267. https://doi.org/10.1016/j.gloenvcha.2006.04.002

38. Burns M, Audouin M, Weaver A. Advancing sustainability science in South Africa. S Afr J Sci. 2006;102:379–384.

39. Mol APJ. Ecological modernization and the global economy. Glob Environ Politics. 2002;2(2):92–115. https://doi.org/10.1162/15263800260047844

40. Helbing D. Globally networked risks and how to respond. Nature. 2013;497:51–59. https://doi.org/10.1038/nature12047

41. Perrow C. Normal accidents: Living with high-risk technologies. Princeton, NJ: Princeton University Press; 1999.

42. Vegter JR. Verslag oor grondwaterondersoek Lichtenburg [Report of a groundwater investigation]. Report no. GH 1116. Pretoria: Department of Water and Sanitation; 1960. Afrikaans.

43. Vahrmeijer JT, Annandale JG, Bristow KL, Steyn JM, Holland M. Drought as a catalyst for change: A case study of the Steenkoppies Dolomite Compartment. In: Schwabe K, Albiac-Murill, J, Connor JD, Hassan R, Meza González L, editors. Drought in arid and semi-arid regions. Dordrecht: Springer; 2013. https://doi.org/10.1007/978-94-007-6636-5_14

44. Holland M. Geohydrology guideline development: Activity 25: Geohydrological assessment of the Steenkoppies Dolomite Compartment. Report prepared for the Department of Water Affairs by Water Geosciences Consulting as part of DWA project number 14/14/5/2: Implementation of dolomite guideline. Pretoria: Department of Water Affairs; 2009.

45. RSA. National development plan 2030 – Our future – Make it work. Pretoria: National Planning Commission; 2012.

46. Aarnoudse E, Bluemling B, Wester P, Qu W. The role of collective groundwater institutions in the implementation of direct groundwater regulation measures in Minqin County, China. Hydrogeol J. 2012;20:1213–1221. https://doi.org/10.1007/s10040-012-0873-z

47. Stockholm Resilience Centre. Governance of social-ecological systems in an increasingly uncertain world needs to be collaborative, flexible and learning-based. Stockholm Resilience Centre Research Insight No. 3: Adaptive governance. Stockholm: Stockholm University; 2012.

Farmer groups and inorganic fertiliser use among smallholders in rural South Africa

AUTHORS:
Sikhulumile Sinyolo[1]
Maxwell Mudhara[2]

AFFILIATIONS:
[1]Economic Performance and Development, Human Sciences Research Council, Pretoria, South Africa
[2]Discipline of Agricultural Economics, University of KwaZulu-Natal, Pietermaritzburg, South Africa

CORRESPONDENCE TO:
Sikhulumile Sinyolo

EMAIL:
sksinyolo@gmail.com

DATES:
Received: 20 Mar. 2017
Revised: 09 Jan. 2018
Accepted: 15 Jan. 2018
Published: 30 May 2018

KEYWORDS:
collective action; smallholder farming; propensity score matching; technology adoption; KwaZulu-Natal

HOW TO CITE:
Sinyolo S, Mudhara M. Farmer groups and inorganic fertiliser use among smallholders in rural South Africa. S Afr J Sci. 2018;114(5/6), Art. #2017-0083, 9 pages. http://dx.doi.org/10.17159/sajs.2018/20170083

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

Smallholder farmers in developing countries are characterised by low uptake of improved farm inputs and weak links to markets. Among other reasons, the high transaction costs that these smallholder farmers incur, as a result of their location in remote areas, inadequate information and missing credit markets, inhibit them from participating in both input and output markets. Organising farmers into groups has been suggested as a potential mechanism for reducing transaction costs. Accordingly, farmer groups have been preferred channels for smallholder farmer support in South Africa, both by the government and donors. However, the impact of these groups on smallholder outcomes such as technology adoption is largely unknown. We investigated the extent to which membership in farmer groups influences the use of improved farm inputs such as inorganic fertiliser among smallholder farmers in South Africa. A sample of 984 households was analysed using the propensity score matching method. Group membership was found to play a positive role in inorganic fertiliser use with a 14% higher chance of inorganic fertiliser use among group members. Among fertiliser users, group members used 170 kg more inorganic fertiliser than did non-members. Further analysis indicated that the effect of group membership on inorganic fertiliser use was heterogeneous among group members. The results suggest that farmer groups play a positive role in the use of improved farm inputs in South Africa. For greater effectiveness of group membership, policymakers should target the less educated, increase the assets of the poor and improve access to extension and information.

Significance:

- The impact of farmer groups on smallholder outcomes such as technology adoption is largely unknown.
- Farmer groups were found to play a positive role in the adoption of agricultural technologies such as inorganic fertilisers.
- Effect of group membership on inorganic fertiliser adoption was heterogeneous among group members.
- Variables that should be targeted for greater effectiveness of collective action were identified.

Introduction

Smallholder farming plays an important role in the livelihoods of the poor in the developing world, in general, and in sub-Saharan Africa in particular. However, smallholder farmers face several constraints that have limited the effectiveness of their farming activities in alleviating the rural poverty and food insecurity challenges. They are often located in remote areas with poor infrastructure, inadequate information and imperfect or missing credit markets, which results in higher transaction costs.[1-5] The higher transaction costs reduce their incentives for participation in both agricultural output and input markets. Moreover, these farmers are poorly endowed with assets and lack adequate access to government support services such as extension and training, which are important in alleviating the effects of high transaction costs.[6-8]

Thus, the smallholder farmers in sub-Saharan Africa are characterised by low uptake of improved farm inputs and weak links to output markets.[9-12] For example, whereas the average intensity of inorganic fertiliser use in Latin America and Asia is about 100 kg/ha, it is below 10 kg/ha in sub-Saharan Africa.[9,13,14] In South Africa, the average inorganic fertiliser application rates of smallholders are significantly below the recommended levels for the respective agro-ecological regions in the country, and are too low and ineffective to sustain crop and soil fertility.[15,16] There is therefore a need for mechanisms to address the challenges that smallholders face to enhance their market participation as well as increase their modern technology adoption rates to ensure that they benefit from these technological advancements.

Several studies have indicated that organising farmers into groups can play a significant role in reducing transaction costs and increasing farmers' market participation and input use.[2,4,5,7,8,17] Farmer groups can provide a variety of services that are key for market access, input use and improved welfare.[4] For instance, buying inputs or selling outputs collectively results in economies of scale, which reduces transportation and transaction costs and increases bargaining power (resulting in favourable prices).[17] Governments and development agencies all over the developing world are placing considerable emphasis on using collective action as a means of effectively linking smallholders with input and output markets.[4,18-22]

The South African government has also been actively promoting collective action through groups among the smallholder farmers.[15,23,24] For example, Output 5 of Outcome 7 of the government's outcomes approach aimed to, among other targets, have at least 30% of smallholder farmers organised into producers' associations or marketing co-operatives by 2014.[24] These farmer groups are expected to give collective power in negotiations for inputs and marketing, thus enhancing the institutional environment for poverty reduction and sustainable and inclusive growth in the rural areas.[24] Farmer groups are also the preferred channel through which most non-governmental organisations (NGOs) and donors reach and support the poor with their food security and poverty

reduction interventions in the rural areas. Hence, several farmer groups and cooperatives have been formed in the smallholder sector. While some of the groups focus on one purpose (e.g. marketing), most of these groups are multi-purpose: helping the farmers access information, secure inputs as well as sell their produce.[15]

Membership in these groups enables pooling of resources, sharing of information as well as collective bargaining, thereby increasing the participation of the smallholder farmers in both the input and output markets. The government as well as NGOs often give free and/or subsidised inputs, credit or training through these farmer groups. While organising farmers into groups may make service provision by government and NGOs easier in many respects, its impact on the farmers themselves is mixed.[4,25] Literature is available on the impact of farmer groups in the output markets, and the results, although somewhat mixed, suggest a positive role of collective action on output market participation.[2,21,26-28] However, few studies, such as Abebaw and Haile[7] have focused on the groups' potential role in improving the adoption of modern technologies. While studies such as Abebaw and Haile[7] are relevant, these studies should not be generalised because technology adoption is context specific.[29,30]

In this study, we investigated the extent to which membership in farmer groups influences inorganic fertiliser use among smallholder farmers in the KwaZulu-Natal (KZN) Province of South Africa, using the propensity score matching technique. This study contributes to the literature in three ways. First, while previous technology adoption studies either failed to include group membership in their models, or did not control for its potential endogeneity in the models, as noted in Abebaw and Haile[7], we controlled for potential endogeneity issues using propensity score matching. Propensity score matching pairs group members and non-members who have similar observable characteristics to control for endogeneity problems that arise from observable variables. Second, we did not assume that group membership has homogenous effects, but also investigated the heterogeneous effects of group membership on fertiliser use. That is, we asked: who is likely to benefit more from being a member of a group? This aspect is important for evidence-based policy and for better targeting of interventions meant to increase modern technology use among smallholder farmers.

Thirdly, we focused on smallholder farmers in KZN. To our knowledge, few studies on this subject, if any, have been done in South Africa in general and KZN in particular. We focused on KZN because smallholder farming is very important in the province, as most rural-based people are employed or self-employed in this sector. According to StatsSA[31], more than 28% of the households in KZN are directly involved in agriculture. The high unemployment rate (33%) in the province, especially among the youth (42%), has resulted in a number of household members who fail to secure employment in urban areas, returning to rural areas to engage in smallholder farming, among other economic activities.[31] The other motivation for selecting KZN as the study area was that, even though it has a huge potential in agriculture owing to good, reliable rainfall (more than 1 000 mm a year), smallholder farmers have failed to tap into this potential because of, among other reasons, their use of rudimentary and out-of-date farming methods.[32] According to KZN-DAE[32], the agricultural output could be increased significantly if modern farming practices were adopted and the natural resources optimally managed for agriculture.

Research methodology

Theoretical framework and selection of variables

Group membership was analysed as a choice problem within a random utility framework[33], following previous literature[2,7,17,34]. According to the random utility theory, a farmer decides to be a group member if the expected utility from group membership (U_i^M) is greater than that of non-membership (U_i), i.e. a farmer chooses group membership if the expected net utility ($U_i^M - U_i^N$) is greater than zero. The unobserved net utility can be expressed as a function of observable elements in the following latent variable model:

$$U_i^* = \beta Z_i + \varepsilon_i, U_i = 1 \text{ if } U_i^* > 0 \quad \text{Equation 1}$$

where U_i is a binary indicator variable that equals 1 for household i in case of group membership and 0 otherwise; β is a vector of parameters to be estimated; Z_i is a vector of household and farm characteristics; and ε_i is an error term.

Group membership is associated with potential costs (membership fees, time, etc.) and benefits (better access to information, inputs, collective bargaining, etc.), which may be perceived differently by different households.[2,17] Individual comparative advantage plays an important role in the choice of whether or not to join a group.[2] Farmers incur different transaction costs because of heterogeneities in household resource endowments and access to information, which results in different market behaviour.[1,2,35,36] The selection of variables into the group membership decision model was based on previous literature.[2,7,26,27,37] Table 1 presents the variables that were considered. These variables include personal details of household head and household characteristics (age, gender, education level, employment status, etc.), wealth and asset endowment (land size, livestock size, income, asset values, etc.), access to support services (extension, credit, training, information, etc.), infrastructural and/institutional support (irrigation, distance to the nearest all-weather road, location/district, etc.).

Table 1: Variables and their descriptions

Variable code	Variable name and description
Outcome variable	
FERTUSE	Fertiliser adoption (1=adopter, 0=non-adopter)
FERTKG	Amount of fertiliser used (kg)
Treatment variable	
GROUP	Farmer group membership (1=yes, 0=no)
Independent variable	
AGE	Age (years)
GENDER	Gender (1=male, 0=female)
MARRIED	Marital status (1=married, 0=unmarried)
EDUCAT	Education level (years)
HHSIZE	Household size (number of people)
LAND	Land size (ha)
TLU	Livestock size (TLUs)
ASSETS	Value of household assets (ZAR)
TOTINC	Annual total household income (ZAR)
EXTENSION	Access to extension (1=yes, 0=no)
INFORM	Number of information sources
CREDIT	Access to credit (1=yes, 0=no)
TRAINING	Access to agricultural training (1=yes, 0=no)
ROADDIST	Distance to the nearest all-weather road (km)
FARMEXP	Farming experience (years)
IRRIGAT	Access to water for irrigation purposes (1=yes, 0=no)
EMPLOYED	Off-farm employment (1=yes, 0=no)
BUSINESS	Ownership of small non-farm business (1=yes, 0=no)
HGWALA	District 1 (1=Harry Gwala, 0=otherwise)
UMZINYAT	District 2 (1=Umzinyathi, 0=otherwise)
UTHUKELA	District 3 (1=Uthukela, 0=otherwise)
UMKHANYA	District 4 (1=Umkhanyakude, 0=otherwise)

Personal characteristics of the household heads such as age, gender, marital status and education level determine group membership by influencing the opportunity cost of time and attitudes towards collective action.[2,7,38] The outcome variables considered were the decision to use inorganic fertilisers and the amount used. The amount of fertiliser used applies to the adopters only, as the non-adopters do not have these figures.

Data

Data were collected from 984 farming households drawn from four districts in KZN. The survey was conducted using a multistage sampling technique. First, 4 districts were purposely chosen out of the 11 districts in the province. The four districts selected were Harry Gwala, Umzinyathi, Umkhanyakude and Uthukela. These districts have a significant number of rural households engaged in farming activities. Second, one local municipality was randomly selected for each district: the Ubuhlebezwe local municipality in the Harry Gwala district; the Msinga local municipality in the Umzinyathi district; the Jozini local municipality in the Umkhanyakude district; and the Imbabazane local municipality in the Uthukela district.

Third, a total of 984 rural households were randomly selected from the four local municipalities. The list of farming households was obtained from the respective local offices of KZN's Department of Agriculture. No stratification was done according to group membership (or any other variable), giving an equal chance for both group members and non-members to be included. The total sample comprised 411 households from Ubuhlebezwe, 239 from Msinga, 143 from Jozini and 191 from Imbabazane. The number of households sampled was not proportional to the population sizes of the respective local municipalities, but was proportional to the number of farming households, as received from the local Department of Agriculture.

The data were collected during the months of October and November 2014 using a structured questionnaire. The questionnaire was administered by experienced enumerators who spoke the local IsiZulu language. These enumerators were trained before the survey. Questionnaire pre-testing, involving 15 rural households, was also done. The ambiguities or difficulties with regard to question wording were noted and remedied during questionnaire pre-testing. The questionnaire included household demographics and socio-economic characteristics; income and wealth endowment; institutional support services; membership in farmer organisations; and inorganic fertiliser use. The data on the use of chemical fertilisers refer to agricultural season prior to the survey, i.e. the 2013/2014 season. All procedures performed in the study were approved by the Human and Social Sciences Research Ethics Committee of the University of KwaZulu-Natal (reference number HSS/0027/015D). Informed consent of the study subjects was obtained and the principles of the 1964 Helsinki declaration were adhered to.

Propensity score matching

To investigate the impact of group membership on inorganic fertiliser adoption, we used the propensity score matching (PSM) method. This non-experimental approach is adopted because group membership is non-random, as an individual household decides to join a group voluntarily. As such, households which are group members might systematically differ from non-members in several socio-economic observable characteristics that may have a direct effect on inorganic fertiliser use. To the extent that group and non-members are different, simply computing the difference between the mean values of the outcome variable of interest of the two categories gives biased results. PSM identifies non-members of groups that are similar to members in their observable characteristics, and comparisons are made in the region of common support. Compared to estimates based on full samples, the impact estimates based on matched samples are less biased and more reliable.[39,40]

To apply PSM in this study, we assumed that our outcome of interest is the amount of inorganic fertiliser used. Further, we assumed that the amount of inorganic fertiliser used by group member *i* is Y_{1i}. The amount of inorganic fertiliser used by a non-member is then assumed to be Y_{0i}. G_i denotes group membership by household *i*, which can take two values: namely $G_i = 1$ if the household is a group member and $G_i = 0$ if the household is a non-member. Our interest was to evaluate the impact of group membership on those households that are group members. The focus is on estimating the average treatment effect on the treated (ATT), the expected treatment effect over the sample of group members, which is estimated as follows:

$$ATT = E[\Delta_i | G_i=1] = E[Y_{1i,t} | G_i=1] - E[Y_{0i,t} | G_i=1] \quad \text{Equation 2}$$

where $E[\Delta | G_i=1]$ is the expected treatment effect; $E[Y_{i1} | G_i=1]$ is the amount of inorganic fertiliser used by the group members; and $E[Y_{i0} | G_i=1]$ is the amount of inorganic fertiliser that would have been used by group members had they not been group members. The ATT captures the change in the amount of fertiliser (outcome) realised by households which are group members subject to their group membership status.

The fundamental evaluation problem is that of missing data.[41] This is because the amount of inorganic fertiliser for the group members, had they not been group members, cannot be observed. Similarly, the amount of inorganic fertiliser used by the non-member households, had they been group members, cannot be observed. In other words, the treatment indicator takes either one or zero, but not both. The PSM procedure was used to generate the missing data.[42-46] PSM can estimate the causal group membership impact as the difference between the amount of fertiliser used for the group members and what would have been the case if these members had not joined groups. Estimating the propensity score, which is simply the probability that a household is a group member, is a crucial step in using matching as an evaluation strategy. The logit model was used to generate the propensity scores as follows:

$$p(Z_i) = Prob(G_i=1 | Z_i) \quad \text{Equation 3}$$

where $p(Z_i)$ is the propensity score; G_i is group membership; and Z_i are the observed household socio-economic characteristics affecting group membership.

According to Becker and Ichino[47], an estimate of the propensity score is not enough to estimate the average treatment effect on the treated (ATT), because the probability of observing two units with exactly the same value of the propensity score is, in principle, zero. Various matching algorithms have been proposed in the literature to determine the region of common support. The most widely used are the nearest-neighbour matching, radius matching, Kernel matching and stratification matching.[41,47-49]

A matching estimator is considered good if, on the one hand, it does not eliminate too many of the original observations from the final analysis, while, on the other hand, it yields statistically equal covariate means for households in the treatment and control groups.[7,50] The nearest-neighbour matching and Kernel matching are reported in this study. The nearest neighbour was chosen because it is generally used in practice because of its ease of implementation, while Kernel matching is a recently developed technique that is gaining popularity in non-experimental literature.[41] The balancing property was selected in estimating the propensity scores. The use of the balancing property ensures that a comparison group is constructed with observable characteristics distributed equivalently across quintiles in both the treatment and comparison groups.[41]

In constructing the matching estimates, the common support was imposed. The treatment observations with weak common support were dropped, as inferences can be made about causality only in the area of common support.[48] All the standard errors were bootstrapped with 1000 repetitions, as suggested by Smith and Todd[41]. The sensitivity of the estimated average adoption effects to hidden bias was tested using the Rosenbaum bounds sensitivity test.[51] This test indicates how strongly an unobservable variable must influence the selection process to undermine or reverse the findings based on matching on observables.[51-53] Previous

studies on group membership impacts such as Abebaw and Haile[7], Cunguara and Darnhofer[54] and Tilahun et al.[34] have used the same approach to test for hidden bias in impact estimates. We focused on positive self-selection in search for evidence for overestimation of ATT, even though the Rosenbaum procedure provides bounds to both positive and negative self-selection.

The estimation of ATT assumes a homogenous treatment effect among the group members. However, as explained in previous studies,[7,40,54,55] the treatment effects are not the same for all the different socio-economic groups within the same treatment group. To investigate the extent to which the treatment effect on fertiliser adoption varies within group members, ordinary least squares regression of the household-level treatment effect on some background characteristics of the group members was estimated.

Results and discussion

Descriptive statistics

Table 2 presents the descriptive statistics of the interviewed households according to group membership status. The table shows that 414 of the sampled households were group members, representing over 42% of the sample. Discussion with the group members indicated that most of these groups have not yet been formally registered, although there is a government effort to ensure that these groups are formally registered as cooperatives. The group members indicated that the groups render several services to their members, such as dissemination of price or market information, collective purchasing of input, output market access, credit and savings, training and information/experience sharing. Some of the benefits of group membership, according to the farmers, is that groups make it easier to access government or NGO support, as these bodies prefer to disseminate extension information, inputs and other forms of support to groups over individuals. The high proportion of smallholders who are members of farmer groups implies that the government's drive to ensure that at least 30% of smallholders are members of groups may have been achieved in the study area. However, the process of formal registration is ongoing.

Table 2 shows that group members were more educated, had bigger households and were wealthier (in terms of land, livestock, assets and income) than the non-members. Table 2 also suggests that group members have better access to support services such as extension, information and credit. The non-farm business owners and those with more farming experience were less likely to be group members. Most sampled households in Uthukela and Umkhanyakude districts were members of the various farmer groups.

Table 2 also shows that inorganic fertiliser use is significantly different between group members (59%) and non-members (53%). Among the fertiliser users, the table indicates that the group members applied nearly

Table 2: Descriptive statistics of sample households according to group membership status

Variable	Means			*t*-test (X^2 tests)
	Pooled sample (*n*=984)	Group members (*n*=414)	Non-members (*n*=570)	
FERTUSE	0.56	0.59	0.53	3.86**
FERTKG[a]	245.40	357.21	172.85	4.37***
AGE	56.11	56.31	55.96	0.42
GENDER	0.47	0.50	0.44	4.46**
EDUCAT	4.67	4.95	4.47	1.78*
HHSIZE	7.04	7.70	6.56	4.98***
LAND	1.93	2.50	1.52	3.44***
TLU	3.53	4.95	2.49	2.18**
ASSETS	82 105.38	88 178.31	77 694.52	4.20***
TOTINC	46 757.43	51 581.08	43 253.93	3.97***
EXTENSION	0.57	0.68	0.49	35.38***
INFORM	2.28	2.65	2.01	9.37***
CREDIT	0.36	0.40	0.32	6.79***
TRAINING	0.41	0.57	0.30	76.63
ROADDIST	17.75	17.28	18.01	-0.31
FARMEXP	18.70	16.25	20.47	4.98***
IRRIGAT	0.46	0.48	0.45	0.86
EMPLOYED	0.20	0.19	0.21	0.22
BUSINESS	0.08	0.11	0.06	5.97**
HGWALA	0.42	0.17	0.60	180.60
UMZINYAT	0.24	0.26	0.23	0.94
UTHUKELA	0.19	0.28	0.13	35.79***
UMKHANYA	0.15	0.29	0.04	120.20***

[a]*Indicates the 554 farmers who used inorganic fertilisers.*

*Significant at the *10%, **5% or ***1% level.*

double the fertiliser quantities applied by the non-members. The box-and-whisker plots in Figure 1 also show that group members used higher quantities of fertiliser than non-members, and that there were relatively higher variations in the fertiliser quantities used among group members than among non-group members. Figure 2 presents the plots showing the quantity of fertiliser used by farmers according to group status and district. Again, the figure shows higher levels of fertiliser use and variations among group members across all districts, with the exception of the Umzinyathi district. However, at this stage, the results should not be used to make inferences regarding the impacts of groups on improved agricultural technology adoption as confounding factors have not been controlled for, which is done through the econometric model later.

Determinants of group membership and estimation of the propensity scores

The logit model was estimated to investigate the factors associated with membership in farmer groups and compute the propensity scores. The results are presented in Table 3. The estimated model fits the data reasonably well as the likelihood ratio X^2 is statistically significant at the 1% level and the model correctly predicts 80% of the sample observations. Most of the variables in Table 3 have the expected signs. The results indicate that group membership is significantly associated with a household's demographic and socio-economic characteristics as well as access to support services.

Table 3 shows that, consistent with previous literature[2,7,27], age is associated with increasing chances of group membership. An additional year is associated with an increase of 0.4% in the likelihood of group membership. This implies that older farmers are more inclined towards working as groups while the younger prefer working as individuals. The reason is that older farmers would have developed more contacts and trust, and have more positive attitudes to group membership than younger farmers. The insignificant coefficient of gender suggests no bias in group membership. This is in contrast to past studies[2,7] which have indicated that female-headed households are less likely to join groups

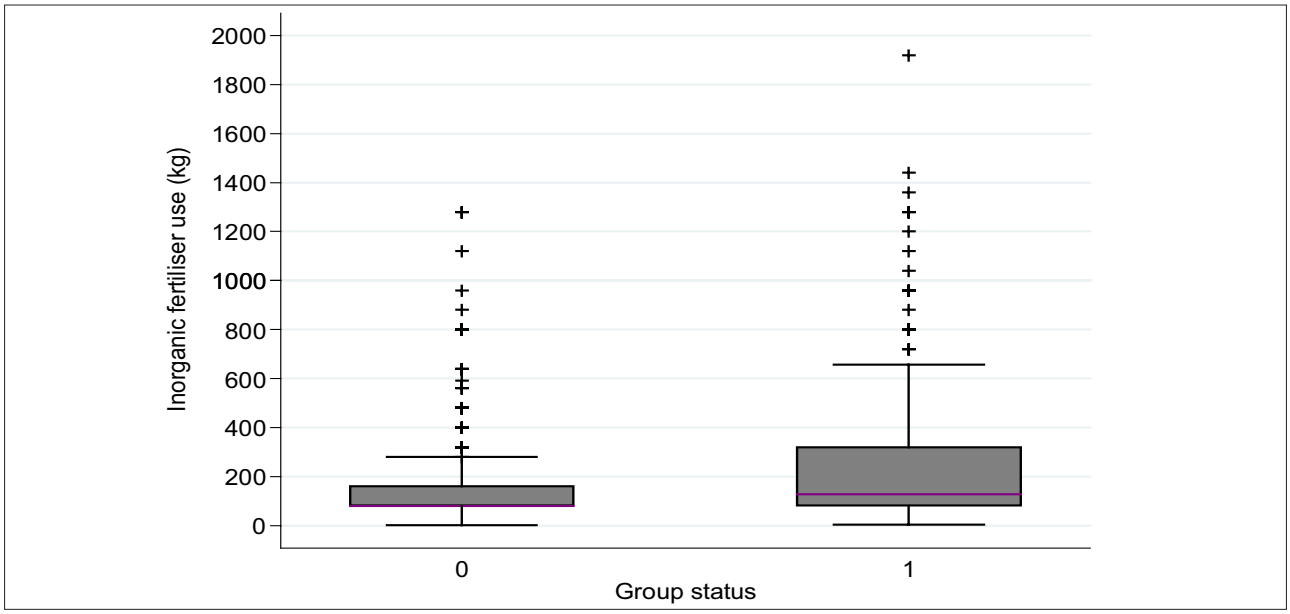


Figure 1: Box-and-whisker plots showing the quantity of fertiliser used by farmers according to group status: 0 indicates non-members and 1 indicates group members.

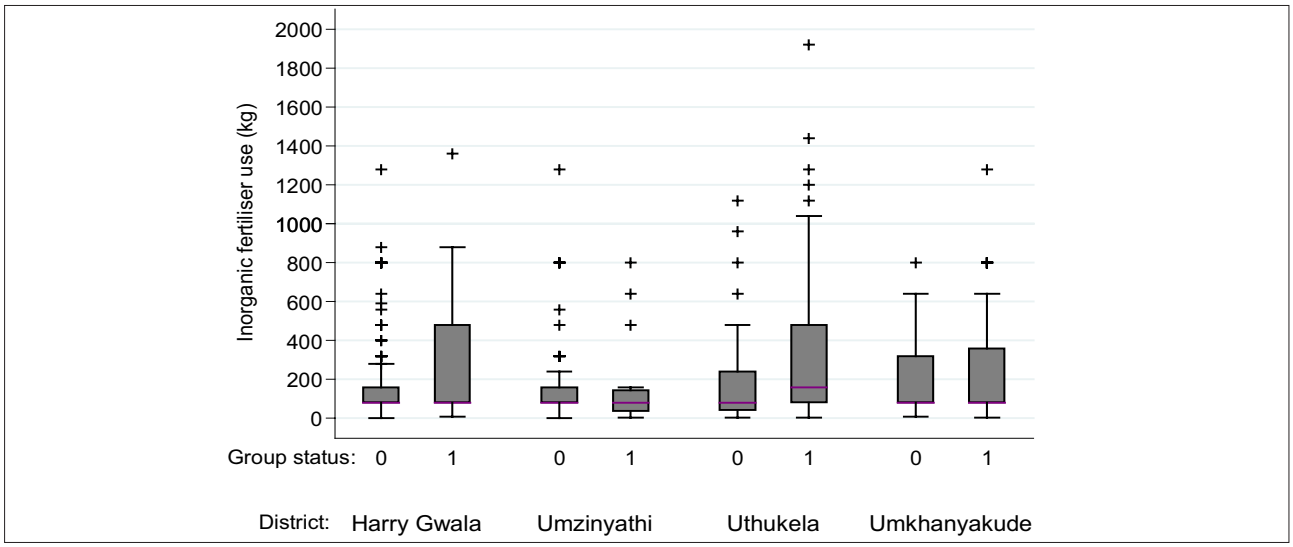


Figure 2: Box-and-whisker plots showing the quantity of fertiliser used by farmers according to district and group status: 0 indicates non-members and 1 indicates group members.

Table 3: Factors determining farmer group membership: logit model results

Variable	Coefficients		Marginal effects	
	Value	Standard error	Value	Standard error
AGE	0.028***	0.008	0.004***	0.001
GENDER	-0.169	0.196	-0.025	0.029
MARRIED	0.269	0.184	0.040	0.028
HHSIZE	0.045*	0.025	0.007*	0.004
EDUCAT	0.039*	0.022	0.006*	0.003
LAND	0.364***	0.085	0.055***	0.012
TLU	-0.014	0.014	-0.002	0.002
ASSETS[a]	0.224*	0.126	0.033*	0.019
INCOME[a]	0.267*	0.149	0.040*	0.022
EXTENSION	0.275**	0.113	0.041**	0.020
INFORM	0.375***	0.082	0.056***	0.012
CREDIT	0.204	0.176	0.030	0.026
TRAINING	1.067***	0.186	0.159***	0.026
ROADDIST	-0.010***	0.002	-0.002***	0.000
FARMEXP	-0.026***	0.007	-0.004***	0.001
IRRIGAT	0.115***	0.016	0.017***	0.006
EMPLOYED	-0.208	0.243	-0.031	0.036
BUSINESS	0.921***	0.334	0.138***	0.049
HGWALA	-2.063***	0.251	-0.308***	0.032
UTHUKELA	0.025	0.269	0.004	0.040
UMKHANYA	1.789***	0.352	0.267***	0.049
_CONSTANT	-8.119	1.922		
Pseudo R^2	0.332			
Likelihood ratio X^2	249.59***			
% predicted correctly	0.80			

[a]*Assets and income values were logged.*

*Significant at the *10%, **5% or ***1% level.*

than male-headed households because women face higher opportunity costs of time as a result of family responsibilities in addition to farming, reducing their incentives for group membership.

In line with studies such as Bernard and Spielman's[26] and Fischer and Qaim's[2], the results indicate that household size is positively associated with group membership. Presumably, bigger households are more likely to participate in groups as a consequence of labour availability. Education, as a proxy of human capital, is also positively associated with participating in groups because the more educated are more likely to understand and interpret information better, which will result in them facing lower transaction costs and benefiting more from the group membership. Table 3 shows that increasing land size is positively correlated with membership in farmer groups. An increase of 1 ha is associated with an increased chance of group membership of 5.5%. The net benefits of farmer group membership increase with increasing farm size, possibly because bigger farms signify increased agricultural production potential. As membership costs are usually fixed, farmers who produce more are likely to benefit more from the groups.

The same pattern also applies to other proxies of physical and financial capital such as asset values, income, irrigation access and ownership of non-farm micro-businesses. The positive relationship between physical as well as financial capital and group membership has been shown by several past empirical studies.[2,7,17,26,27,34,56] The reason is that gains from participation in farmer groups are larger if a household owns complementary assets that enhance successful cooperation.

Access to support services such as extension, information services and training are associated with increased likelihood of group membership. Such services ease access to relevant information about the benefits of group membership. This is in line with previous literature.[7,34] In South Africa, extension officers have been in the forefront of promoting group formation as the government prefers working with farmer groups. As such, extension officers are likely to influence the farmers they contact to form groups. Training also increases the chance of group membership by close to 16%. A counterintuitive result in Table 3 is that distance to the nearest all-weather road has a significant and negative effect on group membership. One would have expected that farmers furthest from all-

weather roads are more likely to join groups to alleviate transportation costs. A plausible explanation is that farmers in isolated areas do not have access to enough information about the benefits of group membership. This is unfortunate, as these are the farmers who would benefit the most from group membership (Table 7).

Farming experience is associated with decreasing chances of group membership. This suggests that experienced farmers would have developed enough individual capacity such that they prefer to work as individuals. The results also show location effects as district dummies, which were included to account for unobserved agro-climatic, institutional, market access and socioeconomic heterogeneities among the sample districts, were significant. In comparison to farmers located in the Umzinyathi district, farmers in Harry Gwala were less likely to participate in groups, while those in Umkhanyakude were more likely to participate. In summary, the logit results show that group participation was biased towards the educated, the relatively wealthier households and households with access to support services such as extension, training and information. Previous studies have also reported that the poor and uneducated tend to be excluded from membership in farmer groups.[7,26]

Impact of group membership on inorganic fertiliser use

The PSM method was employed to estimate the impact of group membership on the probability and level of inorganic fertiliser use. Table 4 shows the impact of group membership on fertiliser use probability. The impacts are estimated using both nearest-neighbour and kernel matching to ensure robustness.

Table 4: Impact of group membership on the probability of inorganic fertiliser use

Matching method	Number of households		ATT	*t*-test
	Treatment	Control		
Nearest neighbour	414	158	0.140 (0.063)	2.210**
Kernel matching	414	461	0.148 (0.052)	2.849***

*Significant at the **5% or ***1% level.*

Table 4 shows that both the matching estimators yield similar results and that group membership has a positive and statistically significant effect on the probability of inorganic fertiliser use. The results indicate that the fertiliser use rate would be 14% lower if the farmers had not participated in farmer groups. The Rosenbaum bounds sensitivity analysis showed that the conclusion would change at a bounds statistic (Γ) of 2.35. This implies that the results are only sensitive to a hidden bias that would more than double the odds of being a group member. Therefore, it is concluded that the results are not very sensitive to hidden bias, because it would require more than 130% of bias to reverse the conclusion.

Table 5 shows the impact of group membership on fertiliser use level. The table shows that the fertiliser use level would have been between 134 kg and 168 kg lower had farmers not joined farmer groups, indicating the positive role that farmer groups play in fertiliser use. Rosenbaum bounds test showed that the conclusion would change at $\Gamma = 4.35$, implying that the results are not very sensitive to hidden bias because it would require more than 300% of bias to reverse the conclusion.

Table 5: Impact of farmer group membership on inorganic fertiliser use level

Matching method	Number of households		ATT	*t*-test
	Treatment	Control		
Nearest neighbour	218	88	167.41 (52.95)	3.161***
Kernel matching	218	288	134.23 (49.60)	2.706***

*Significant at the ***1% level.*

To further evaluate the reliability of the above reported estimates, the balancing tests based on nearest neighbour were done and the results are presented in Table 6. The table shows that, after matching, both group members and non-members have characteristics that are statistically similar. The test for equality of the two group means shows that there is no statistically significant difference between members and non-members after matching. This contrasts with the unmatched sample presented in Table 2 which indicated statistically significant differences in several covariates between the two groups. The standardised differences (% bias) for the mean values of all the covariates between members and non-members are below 20%, implying that the balancing requirement is adequately satisfied.[57]

Table 6: Test of matching quality

Variable	Means		% Bias	% Reduction |bias|	*p*-value of equality of mean
	Members	Non-members			
AGE	56.31	55.60	5.4	-99.8	0.42
EDUCAT	4.95	5.14	-2.4	68.3	0.50
HHSIZE	7.70	7.71	-4.6	59.6	0.97
LAND	2.50	3.14	-13.5	68.7	0.23
TLU	4.95	4.18	4.1	68.8	0.57
ASSETS[a]	11.25	11.31	-9	66	0.11
TOTINC[a]	10.65	10.60	7.9	46	0.24
EXTENSION	0.68	0.71	-7.5	81	0.26
INFORM	2.65	2.74	-8.7	85.7	0.19
CREDIT	0.40	0.36	9.1	46	0.20
TRAINING	0.57	0.53	10.1	82.6	0.16
ROADDIST	17.28	17.83	-1.4	31.7	0.84
FARMEXP	16.25	15.60	5	84.5	0.43
IRRIGAT	0.48	0.44	7.3	-21.6	0.30
EMPLOYED	0.19	0.16	8.5	-181.2	0.20
BUSINESS	0.11	0.10	2.6	83.2	0.73
HGWALA	0.17	0.17	0	100	1.00
UTHUKELA	0.28	0.30	-4.3	88.9	0.59
UMKHANYA	0.29	0.29	0	100	1.00

[a]*Assets and income values were logged.*

Impact heterogeneity

To investigate the extent to which the treatment effect on fertiliser adoption differs among group members, the ordinary least squares regression was estimated and results are presented in Table 7. The table shows that the impact of group membership is not the same among members. The results show that group membership increases inorganic fertiliser use more for the less educated than for those with more education. This is, as explained by Abebaw and Haile[7], an important result as the poorly educated are less likely to adopt improved farm inputs as a result of their limited ability to understand and interpret new information on technologies. The impact is also larger for those with smaller household sizes and those households with more land and assets and fewer livestock. This result suggests that group membership benefits the richer more than it does the poorer.

Table 7: Heterogeneous fertiliser use impacts among group members

Variable	Coefficient	Standard error
AGE	-0.001	0.002
GENDER	-0.039	0.053
MARRIED	-0.015	0.051
EDUCAT	-0.012*	0.006
HHSIZE	-0.013**	0.006
LAND	0.025***	0.008
TLU	-0.004**	0.002
ASSETS	0.097***	0.036
INCOME	0.052	0.038
EXTENSION	0.090*	0.055
INFORM	0.043*	0.022
CREDIT	0.043	0.048
TRAINING	-0.067	0.048
ROADDIST	0.002***	0.001
FARMEXP	0.006***	0.002
IRRIGAT	0.156***	0.047
EMPLOYED	-0.073	0.073
BUSINESS	0.275***	0.091
HGWALA	0.047	0.076
UTHUKELA	0.344***	0.067
UMKHANYA	-0.117*	0.070
_CONS	-1.335	0.529
N	414	
F	6.16***	
R^2	0.248	

*Significant at the *10%, **5% or ***1% level.*

The farmers with access to extension and information would benefit more from farmer groups, as would farmers with access to irrigation. The results also suggest that the impact of farmer groups on inorganic fertiliser adoption is larger for experienced farmers, farmers located further from all-weather roads and owners of small, non-farm businesses. The greater impact of group membership for the farmers further away from all-weather roads suggests that groups contribute towards alleviating the transaction costs that these isolated farmers face. The fact that gender is not significant suggests that the impact of farmer groups on fertiliser adoption is the same for both male and female farmers, indicating no gender bias.

Conclusions and policy implications

The South African government has identified increased smallholder productivity and commercialisation as an integral part of the strategy for stimulating rural economic development and reducing poverty. The government has also invested considerable effort in organising smallholder farmers into groups to establish an enabling institutional environment for agricultural intensification and achieving sustainable, inclusive and better growth in the sector. Limited research has investigated the role of these farmer groups in improving the adoption of modern farm inputs. We investigated the impact of groups on the use of inorganic fertilisers using cross-sectional data from 984 households and the PSM technique.

The empirical results indicate that participation in farmer groups significantly and positively influences fertiliser use. Group membership improves the average fertiliser application rate by about 14%, and the fertiliser use level by 134–168 kg. The Rosenbaum bounds tests indicate the impact estimates obtained using the PSM approach were robust to hidden bias. The results also show greater group membership impact for the less educated, the wealthier (more land and assets), the irrigators, those with access to extension and information and those located further from all-weather roads. The findings suggest that the government's strategy of organising farmers into groups for improved smallholder production activities should continue, as groups raise the demand for improved farm inputs such as inorganic fertilisers. For greater effectiveness of membership of groups in improving modern technology adoption among smallholders, policymakers should target the less educated, increase the assets of the poor and improve access to extension and information.

Authors' contributions

S.S. was responsible for conceptualisation, methodology, data collection, data analysis, and writing the initial draft and revisions. M.M. was responsible for the methodology and critically reviewing the writing and revisions.

References

1. Key N, Sadoulet E, De Janvry A. Transactions costs and agricultural household supply response. Am J Agric Econ. 2000;82(2):245–259. https://doi.org/10.1111/0002-9092.00022
2. Fischer E, Qaim M. Linking smallholders to markets: Determinants and impacts of farmer collective action in Kenya. World Dev. 2012;40(6):1255–1268. https://doi.org/10.1016/j.worlddev.2011.11.018
3. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, et al. Food security: The challenge of feeding 9 billion people. Science. 2010;327(5967):812–818. https://doi.org/10.1126/science.1185383
4. Hellin J, Lundy M, Meijer M. Farmer organization, collective action and market access in Meso-America. Food Pol. 2009;34:16–22. https://doi.org/10.1016/j.foodpol.2008.10.003
5. Markelova H, Meinzen-Dick R, Hellin J, Dohrn S. Collective action for smallholder market access. Food Pol. 2009;34:1–7. https://doi.org/10.1016/j.foodpol.2008.10.001
6. Alene AD, Manyong VM, Omanya G, Mignouna HD, Bokanga M, Odhiambo G. Smallholder market participation under transactions costs: Maize supply and fertilizer demand in Kenya. Food Pol. 2008;33(4):318–328. https://doi.org/10.1016/j.foodpol.2007.12.001
7. Abebaw D, Haile MG. The impact of cooperatives on agricultural technology adoption: Empirical evidence from Ethiopia. Food Pol. 2013;38:82–91. http://dx.doi.org/10.1016/j.foodpol.2012.10.003
8. Kruijssen F, Keizer M, Giuliani A. Collective action for small-scale producers of agricultural biodiversity products. Food Pol. 2009;34:46–52. https://doi.org/10.1016/j.foodpol.2008.10.008
9. African Union/New Partnership to Africa's Development (AU/NEPAD). The Abuja Declaration on Fertilizers for an African Green Revolution: Status of implementation at regional and national levels. Johannesburg: AU/NEPAD; 2011.
10. Diiro GM, Ker AP, Sam AG. The role of gender in fertiliser adoption in Uganda. Afr J Agric Resour Econ. 2015;10(2):117–130.
11. Yirga C, Hassan RM. Determinants of inorganic fertiliser use in the mixed crop-livestock farming systems of central highlands of Ethiopia. Afr Crop Sci J. 2013;21(3):669–681.
12. Mwangi M, Kariuki S. Factors determining adoption of new agricultural technology by smallholder farmers in developing countries. J Econ Sustain Dev. 2015;6(5):208–216.

13. Crawford EW, Jayne TS, Kelly VA. Alternative approaches for promoting fertilizer use in Africa. Agriculture and Rural Development Discussion Paper 22. Washington DC: The International Bank for Reconstruction and Development/The World Bank; 2006.

14. Morris M, Kelly VA, Kopicki RJ, Byerlee D. Fertilizer use in African agriculture: Lessons learned and good practice guidelines. Washington DC: World Bank; 2007. https://doi.org/10.1596/978-0-8213-6880-0

15. Department of Agriculture, Forestry and Fisheries (DAFF). A framework for the development of smallholder farmers through cooperatives development. Pretoria: DAFF; 2012.

16. Mkhabela TS, Materechera SA. Factors influencing the utilization of cattle and chicken manure for soil fertility management by emergent farmers in the moist Midlands of KwaZulu-Natal Province, South Africa. Nutr Cycl Agroecosys. 2003;65:151–162. https://doi.org/10.1023/A:1022156210667

17. Fischer E, Qaim M. Smallholder farmers and collective action: What determines the intensity of participation? J Agric Econ. 2014;65(3):683–702. https://doi.org/10.1111/1477-9552.12060

18. Kaganzi E, Ferris S, Barham J, Abenakyo A, Sanginga P, Njuki J. Sustaining linkages to high value markets through collective action in Uganda. Food Pol. 2009;34:23–30. https://doi.org/10.1016/j.foodpol.2008.10.004

19. HLPE. Investing in smallholder agriculture for food security. HLPE report 6. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security. Rome: Food and Agricultural Organization (FAO); 2013.

20. IFAD. Promoting market access for the rural poor in order to achieve the Millennium Development Goals: Discussion paper [document on the Internet]. c2003 [cited 2014 May 27]. Available from: http://www.ifad.org/gbdocs/gc/26/e/markets.pdf.

21. Narrod C, Roy D, Okello J, Avenda-o B, Rich K, Thorat A. Public–private partnerships and collective action in high value fruit and vegetable supply chains. Food Pol. 2009;34:8–15. https://doi.org/10.1016/j.foodpol.2008.10.005

22. Liverpool-Tasie LSO. Farmer groups and input access: When membership is not enough. Food Pol. 2014;46:37–49. https://doi.org/10.1016/j.foodpol.2014.01.006

23. Ortmann GF, King RP. Agricultural cooperatives I: History, theory and problems. Agrekon. 2007;46(1):18–46. https://doi.org/10.1080/03031853.2007.9523760

24. RSA. Outcome 7: Vibrant, equitable and sustainable rural communities and food security for all [document on the Internet]. c2010 [cited 2016 Jul 05]. Available from: http://www.gov.za/sites/www.gov.za/files/Delivery%20Agreement%20-%20Outcome%207_0.pdf

25. Cadzow H, Binns T. Are groups a good thing? Evaluating group associations among vegetable farmers in Freetown, Sierra Leone. Dev Pract. 2016;26(4):406–419. https://doi.org/10.1080/09614524.2016.1159662

26. Bernard T, Spielman DJ. Reaching the rural poor through rural producer organizations? A study of agricultural marketing cooperatives in Ethiopia. Food Pol. 2009;34(1):60–69. http://dx.doi.org/10.1016/j.foodpol.2008.08.001

27. Bernard T, Taffesse AS, Gabre-Madhin E. Impact of cooperatives on smallholders' commercialization behavior: Evidence from Ethiopia. Agric Econ. 2008;39(2):147–161. https://doi.org/10.1111/j.1574-0862.2008.00324.x

28. Shiferaw B, Obare G, Muricho G. Rural market imperfections and the role of institutions in collective action to improve markets for the poor. Nat Resour Forum. 2008;32(1):25–38. https://doi.org/10.1111/j.1477-8947.2008.00167.x

29. Kafle B. Determinants of adoption of improved maize varieties in developing countries: A review. Int Res J Appl Basic Sci. 2010;1(1):1–7.

30. Ogada M, Mwabu G, Muchai D. Farm technology adoption in Kenya: A simultaneous estimation of inorganic fertilizer and improved maize variety adoption decisions. Agric Econ. 2014;2(1):1–18. https://doi.org/10.1186/s40100-014-0012-3

31. Statistics South Africa (StatsSA). Census 2011 Municipal report – KwaZulu-Natal [document on the Internet]. c2012 [cited 2015 Mar 09]. Available from: http://www.statssa.gov.za/Census2011/Products/KZN_Municipal_Report.pdf.

32. KwaZulu-Natal Department of Agriculture and Environmental Affairs (KZNDAE). Annual performance plan 2012/13. Pietermaritzburg: KZNDAE; 2012.

33. McFadden D. The measurement of urban travel demand. J Public Econ. 1974;3(4):303–328. http://dx.doi.org/10.1016/0047-2727(74)90003-6

34. Tilahun M, Maertens M, Deckers J, Muys B, Mathijs E. Impact of membership in frankincense cooperative firms on rural income and poverty in Tigray, Northern Ethiopia. For Policy Econ. 2016;62:95–108. http://dx.doi.org/10.1016/j.forpol.2015.08.009

35. Barrett CB. Smallholder market participation: Concepts and evidence from eastern and southern Africa. Food Pol. 2008;33:299–317. https://doi.org/10.1016/j.foodpol.2007.10.005

36. Boughton D, Mather D, Barrett CB, Benfica R, Abdula D, Tschirley D, et al. Market participation by rural households in a low-income country: An asset-based approach applied to Mozambique. Faith Econ. 2007;50:64–101.

37. Wollini M, Zeller M. Do farmers benefit from participating in speciality markets and cooperatives? The case of coffee marketing in Costa Rica. Agric Econ. 2007;37:243–248. https://doi.org/10.1111/j.1574-0862.2007.00270.x

38. Ostrom E. Collective action and the evolution of social norms. J Econ Perspect. 2000;14:137–158. https://doi.org/10.1257/jep.14.3.137

39. Rubin D, Thomas D. Combining propensity score matching with additional adjustments for prognostic covariates. J Am Stat Assoc. 2000;95:573–585. https://doi.org/10.1080/01621459.2000.10474233

40. Abebaw D, Fentie Y, Kassa B. The impact of a food security program on household food consumption in Northwestern Ethiopia: A matching estimator approach. Food Pol. 2010;35(4):286–293. http://dx.doi.org/10.1016/j.foodpol.2010.01.002

41. Smith J, Todd P. Does matching overcome LaLonde's critique of nonexperimental estimators? J Econom. 2005;125(1–2):303–353. https://doi.org/10.1016/j.jeconom.2004.04.011

42. Baker JL. Evaluating the impact of development projects on poverty: A handbook for practitioners. Washington DC: World Bank; 2000. https://doi.org/10.1596/0-8213-4697-0

43. Jalan J, Ravallion M. Estimating the benefit incidence of an antipoverty program by propensity score matching. J Bus Econ Stat. 2003;21(1):19–30. https://doi.org/10.1198/073500102288618720

44. Ravallion M. Evaluating anti-poverty programs. In: Evenson RE, Schultz TP, editors. Handbook of agricultural economics. Amsterdam: Elsevier; 2008.

45. Blundell R, Costa-Dias M. Evaluation methods for non-experimental data. Fisc Stud. 2000;21(4):447–468. https://doi.org/10.1111/j.1475-5890.2000.tb00031.x

46. Khandker SR, Koolwal GB, Samad HA. Handbook on impact evaluation: Quantitative methods and practice. Washington DC: The World Bank; 2010.

47. Becker OS, Ichino A. Estimation of average treatment effects based on propensity scores. Stata J. 2002;2(4):358–377.

48. Heckman JJ, Ichimura H, Todd P. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. Rev Econ Stud. 1997;64(4):605–654. https://doi.org/10.2307/2971733

49. Dehejia R, Wahba S. Propensity score matching methods for nonexperimental causal studies. Rev Econ Stat. 2002;84(1):151–161. https://doi.org/10.1162/003465302317331982

50. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. Discussion paper no. 1588 [document on the Internet]. c2005 [cited 2014 Feb 27]. Available from: http://ftp.iza.org/dp1588.pdf

51. Rosenbaum PR. Observational studies. New York: Springer; 2002. https://doi.org/10.1007/978-1-4757-3692-2_3

52. Rosenbaum PR. Heterogeneity and causality. Am Stat. 2005;59(2):147–152. https://doi.org/10.1198/000313005X42831

53. Rosenbaum PR. Discussing hidden bias in observational studies. Ann Intern Med. 1991;115(11):901–905. https://doi.org/10.7326/0003-4819-115-11-901

54. Cunguara B, Darnhofer I. Assessing the impact of agricultural technologies on household income in rural Mozambique. Food Pol. 2011;36:378–390. https://doi.org/10.1016/j.foodpol.2011.03.002

55. Ali A, Abdulai A. The adoption of genetically modified cotton and poverty reduction in Pakistan. J Agric Econ. 2010;61(1):175-92. https://doi.org/10.1111/j.1477-9552.2009.00227.x

56. Francesconi GN, Heerink N. Ethiopian agricultural cooperatives in an era of global commodity exchange: Does organizational form matter? J Afr Econ. 2011;20:153–177. https://doi.org/10.1093/jae/ejq036

57. Rosenbaum PR, Rubin DB. Constructing control group using a multivariate matched sampling method that incorporates the propensity score. Am Stat. 1985;39:33–38.

Power laws, demography and entrepreneurship in selected South African regions

AUTHOR:
Daan Toerien[1]

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

CORRESPONDENCE TO:
Daan Toerien

EMAIL:
dtoerien@gonet.co.za

DATES:
Received: 01 Aug. 2017
Revised: 26 Oct. 2017
Accepted: 15 Jan. 2018
Published: 30 May 2018

KEYWORDS:
power laws; Zipf's law; demography; entrepreneurship; agglomeration patterns

HOW TO CITE:
Toerien D. Power laws, demography and entrepreneurship in selected South African regions. S Afr J Sci. 2018;114(5/6), Art. #2017-0255, 8 pages. http://dx.doi.org/10.17159/sajs.2018/20170255

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Gouritz Cluster Biosphere Reserve; University of the Free State

Statistically significant Pareto-like log-log rank-size distributions were recorded for population and enterprise agglomeration in the towns of three different regions of South Africa, and are indicative of skewed distributions of population and enterprise numbers in regional towns. There were no distinct differences between groups of towns of regions from different parts of the country. However, the regional agglomerations differed from those of groups of towns randomly selected from a database. Regions, therefore, appear to have some uniqueness regarding such agglomerations. The identification of Zipf-like links between population and enterprise growth in regional towns still does not fully explain why some towns grow large and others stay small and there is a need to further explore these issues. The extreme skewness in population and enterprise numbers of different towns' distributions should, however, be considered in local economic development planning and execution.

Significance:

- This contribution illustrates that the populations and enterprises of South African regional towns are distributed in orderly ways (called Pareto distributions) that result in some being large/many and others small/few.

Introduction

Accounting for the way populations are distributed over different geographical locations and their evolution over time is important. The distribution of populations across geographical areas is not random[1]: there is a strong tendency toward agglomeration, i.e. populations are concentrated within common restricted areas such as cities, which results in a few large cities and many smaller cities. This striking pattern of geographical agglomeration is called Zipf's law for cities.[2] For instance, the size distribution of cities in the USA is startlingly well described by a simple power law,[3] which essentially states that the probability that the size of a city is greater than some S is proportional to 1/S. Zipf's law is a special case of a Pareto distribution.[4]

Although normal (Gaussian) distributions and related quantitative methods are still relevant for a significant portion of organisational research, the increasing discovery of power laws signifies that Pareto rank-frequency distributions are pervasive and indicative of non-linear organisational dynamics.[4] Researchers ignoring Pareto distributions risk drawing false conclusions and promulgating useless advice to practitioners. Most managers face extremes, not averages.[4] Morudu and du Plessis[5] reported Pareto rank-frequency distributions for population, employment and gross economic value addition data of municipalities in South Africa. Their results hint of a limited impact of national policies such as GEAR (the Growth, Employment and Redistribution), ASGISA (Accelerated and Shared Growth Initiative for South Africa) and NGP (the New Growth Path) plans since 2001, and the marginal spatial impact of local economic development plans on economic variables at municipal level.

Patterns of a few large towns and many smaller ones have also been observed in regional studies of South African towns.[6-12] It is unknown if these patterns adhere to a rank-size distribution consistent with Zipf's law. Considering the above warning[4] and the implications reported for municipalities[5], it is necessary to investigate the possible presence of Pareto (Zipf-like) population and/or enterprise distributions in the towns of South African regions. The presence of linear regularities (proportionalities) in the form of statistically significant correlations between the demographic and entrepreneurial characteristics of South African towns[7-19] could signal the possible presence of Pareto rank-size distributions in regions.

Knudsen[20] examined the size distribution of cities in Denmark by way of three questions: (1) Does Zipf's law apply to the population distribution of Danish cities? (2) What are the implications of Zipf's law for models of local growth? (3) Is there a Zipf's law for firms? Knudsen found that Zipf's law applies to Danish cities and that the size pattern of more than 14 000 Danish production companies follows a rank-size distribution consistent with Zipf's law. He did not examine the distribution patterns of the number of enterprises in Danish cities.

Knudsen's approach[20] provided guidance to this investigation about possible Pareto population and/or enterprise distributions in the towns of South African regions, i.e. the use of questions to examine the distribution patterns. The following questions were examined: (1) Does Zipf's law or Pareto rank-size distributions apply to the population distribution of towns in different regions of South Africa? (2) If Zipf's law (or a Pareto rank-size distribution) applies to town populations, does it also apply to the number of enterprises in these towns? (3) If Zipf's law (or a Pareto rank-size distribution) applies to populations and/or enterprises, do different regions differ from one another (in other words, do they have uniqueness)? (4) Should such distributions, if they exist, be considered in local economic development planning and support? The purpose of this contribution is to provide answers to these questions.

Approach used in this contribution

Previously studied towns of three different regions of South Africa were selected for this analysis (Table 1). The first group includes 12 towns of the Eastern Cape Karoo (EC Karoo) that have previously been extensively

studied.[6,9] These studies included an analysis of data covering almost a century (1911 to 2006).[11] The second group includes 29 Karoo towns included in a recent Shale Gas Strategic Environmental Assessment of the Karoo.[21] Further study of these towns could provide information essential for decision-making about shale gas production. The third group includes the towns of the Gouritz Cluster Biosphere Reserve (GCBR) in the southern Cape.[22]

Table 1: The towns of the Eastern Cape Karoo (EC Karoo), the strategic environmental assessment (SEA) study area and the Gouritz Cluster Biosphere Reserve (GCBR). Towns that fall into two of the three regions are indicated in bold.

No.	EC Karoo	SEA study area	GCBR
1	**Aberdeen**	**Aberdeen**	Albertinia
2	**Cradock**	Beaufort-West	Barrydale
3	**Graaff-Reinet**	Burgersdorp	Calitzdorp
4	**Hofmeyer**	Carnarvon	De Rust
5	**Jansenville**	Colesberg	Great Brak River
6	**Middelburg**	**Cradock**	Heidelberg
7	**Pearston**	Fort Beaufort	Ladismith
8	**Somerset East**	Fraserburg	Montagu
9	**Steynsburg**	**Graaff-Reinet**	Mossel Bay
10	Steytlerville	**Hofmeyr**	Oudtshoorn
11	Venterstad	**Jansenville**	**Prince Albert**
12	Willowmore	Klipplaat	Riversdal
13		Lady Frere	Stilbaai
14		Laingsburg	Swellendam
15		Loxton	Uniondale
16		Merweville	
17		**Middelburg**	
18		Murraysburg	
19		Nieu-Bethesda	
20		Noupoort	
21		**Pearston**	
22		**Prince Albert**	
23		Queenstown	
24		Richmond	
25		**Somerset East**	
26		**Steynsburg**	
27		Sutherland	
28		Victoria West	
29		Williston	

Population and enterprise distribution in cities

Why do cities exist, and why do they vary in size? These fundamental questions have received a considerable amount of attention from regional and urban economists in recent years.[23] Cities are thought to arise to give consumers easy access to a large variety of goods or because of the 'external' effects of consumer location or because of the advantages of proximity of consumers to their workplaces. Although these reasons are probably important, Brakman et al.[23] believe they do not explain why cities are spread out unevenly across space or explain why systems of cities exist.

Already in 1682, Alexandre Le Maitre remarked on a systematic pattern in the size distribution of French cities.[1] Power laws were discovered more than a hundred years ago by Vilifredo Pareto.[24] Felix Auerbach, in 1913, and George Kingsley Zipf, in 1949, formally established an empirical regularity: the sizes of the large cities are inversely proportional to their ranks.[1] The proportionality of rank and size implies a power distribution with exponent equal to one – a phenomenon that became known as Zipf's law for cities, a special case of a Pareto distribution. It is a striking pattern of agglomeration that may well be the most accurate regularity in economics and it holds for many countries and dates.[2] If a sample of cities is ranked according to population size and presented as a graph of log population size (independent variable) and log rank (dependent variable), a straight line with slope -1 indicates a Zipf distribution.[23] If the slope is higher than one, cities are more dispersed than predicted by Zipf's law, and if the slope is less than one, cities are more even-sized than the prediction.[23]

The reason for the existence of Zipf's law for population distribution is still rather obscure and Krugman[3] remarked: 'At this point we are in the frustrating position of having a striking empirical regularity with no good theory to account for it.' However, Gabaix[2] stated that the reason why cities become large is essentially because of inertia in the creation of jobs: the number of new jobs is roughly proportional to the number of existing jobs. Eeckhout[1] remarked that once population mobility is understood, the underlying economic mechanisms can be examined. Agglomeration and residential mobility of the population between different geographical locations are tightly connected to economic activity; the evolution of the population across geographical locations is an extremely complex amalgam of incentives and actions taken by many individuals, enterprises and organisations.

A pattern of some large towns combined with a number of smaller towns is also present in the three regions under consideration here. However, it is not known if there are Pareto rank-size regularities as far as populations are concerned, and, if present, whether they adhere to Zipf's law or differ from one another.

What about enterprise development? Axtell[25] mentioned that Gibrat reported a lognormal distribution of the sizes of French industrial firms in 1931. Such a distribution was later also recorded in the UK.[26] Axtell[27] reported that the distribution of US firm sizes closely followed the Pareto distribution with an exponent near unity, i.e. the Zipf distribution. Knudsen[20] reported that the size distribution of Danish production companies had a strong fit with a Pareto rank-size distribution with an exponent of 0.741.

No record was found that the distribution patterns of the number of enterprises in countries or regions have received research attention. Given the statistically significant relationships between population sizes and enterprise numbers frequently recorded for South African towns,[7-19] it is clear that town size (measured by the population number), the creation of jobs, and thus enterprise numbers and entrepreneurial decisions, have similar variance patterns. It is, therefore, also necessary to examine if the enterprise numbers of regions have Pareto rank-size distributions. If they do, such regularities could be used in development decisions and support as suggested by Andriani and McKelvey[4] and used by Morudu and du Plessis[5].

EC Karoo towns

The EC Karoo and its 12 towns (Table 1) provided the focus of a number of studies.[6,9,11] This area is a sub-region of the Karoo, wholly included in the Eastern Cape Province. Its selection for study by Nel and Hill[6] was based on the availability of comparable and continuous census data records over an extended period. They identified the following trends

based on nearly 100 years of continuous records: shifts in agricultural production, small town and rural population change, and evolving small town economies. The 12 towns were also used in a study of enterprise proportionality phenomena in small towns of the EC Karoo.[9] The towns' enterprises were identified, enumerated and classified into different business sectors. Statistical analyses were used to examine the enterprise dynamics. Regional proportionalities, i.e. fairly constant ratios between business sector enterprise numbers and total enterprise numbers in the towns, were used to construct a 'regional enterprise structure'.

The 12 towns have also been used to address the question of whether the proportionality between population numbers and enterprise numbers in South African towns was present at earlier times. Access to the century-long database of Nel and Hill[6] enabled such a study.[11] Ten data sets were extracted in which the years of the population estimates and enterprise counts in a specific data set differed by at most 2 years. Proportionalities were present over the century and a detailed picture of the relationship between population dynamics and enterprise dynamics was developed.[11] The Nel and Hill[6] data sets also lend themselves to examination of the time-dependence of rank-size distributions of the population and enterprise numbers of the EC Karoo. This region was chosen as the first region in the present study.

Karoo towns

The Karoo, which occupies some 40% of the surface area of South Africa, has a continuous census record and a network of small towns of differing sizes, which made it a suitable area to research aspects of small town development.[28] In addition, shale gas development in the Karoo is being considered by the South African government. An area of 171 811 km^2 of the Central Karoo, delimited by the applications for exploration rights for shale gas lodged by different companies, plus a 20-km buffer, constituted the study area of a strategic environmental assessment that considered shale gas development in the Karoo.[21] The study area includes 29 towns (Table 1).

The Karoo was slowly urbanised after 1785 when its first town, Graaff-Reinet, was founded.[29] By 1850 only 10 of the selected 29 towns had been founded, but, with a couple of exceptions, all had been founded by 1900. These towns exhibit a spread of population and enterprise sizes that raises a question about the possible presence of rank-size distributions.

There is an overlap of nine towns between the groups of the EC Karoo and the strategic environmental assessment study area (Table 1). This overlap was considered acceptable because comparison of the results for the two groups would help to determine if specific rank-size distributions, if present, are typical of a specific region. The shale gas development strategic environmental assessment study area, with its 29 towns, was selected as the second region in this study.

The Gouritz Cluster Biosphere Reserve

The GCBR is located in the southern Cape area of South Africa. It is globally unique as it is the only area in the world where three recognised biodiversity hotspots converge: the Fynbos, Succulent Karoo and Maputaland-Tongoland-Albany hotspots.[22] Two mountain ranges (the Swartberg Mountains in the north and the Langeberg/Outeniqua Mountains in the south) separate the GCBR into two separate geographical sub-regions. To the north and nestled between the Swartberg and the Langeberg/Outeniqua/Tsitsikamma mountain ranges lies a semi-arid to arid valley, the Little Karoo.[30] In the south lies a more verdant coastal plain bordered by the Langeberg mountains in the north and the Indian Ocean in the south.

Urbanisation of the GCBR was also slow. Only one town, Swellendam, had been founded by 1800 for administrative purposes.[29] By 1850 there were only five towns but thereafter the pace of urbanisation increased to the extent that all 15 of the GCBR towns (Table 1) had been founded by 1900. Most of these towns were founded to cater for a rural population's needs of religious services and not for commercial reasons.[29] The GCBR was selected as the third region of this study.

Methods

Population size rank-size analyses

Population numbers for 1946 were obtained from a government report which provides information from 1904 to 1970.[31] Population numbers for 2001 and 2011 of the towns were sourced from a German website.[32] Population estimates for 2014 or 2016 were based on 2011 data extended by the growth rate between 2001 and 2011.

The towns of each of the regions for a specific year were ranked from highest to lowest according to their population numbers. The following regression was then calculated for each time period and region:

$$\text{Log rank}_{1,2\ldots n} = a - b(\text{log population size}_{1,2\ldots n}) \qquad \text{Equation 1}$$

where a is the intercept, b is the regression coefficient and n is the number of towns in a region. Microsoft Excel software was used for the calculations.

Enterprise numbers rank-size analyses

Enterprise numbers for 1946/1947 and 2013/2014 or 2015/2016 of towns of all three regions were determined according to the methods of Toerien and Seaman[7], using telephone directories for the specific year.

The towns of the different regions for a specific year were ranked from highest to lowest according to their enterprise numbers. The following regression was then calculated for each time period and region:

$$\text{Log rank}_{1,2\ldots n} = a - b(\text{log enterprise numbers}_{1,2\ldots n}) \qquad \text{Equation 2}$$

where a is the intercept, b is the regression coefficient and n is the number of towns in a region. Microsoft Excel software was used for the calculations.

Time dependence of rank-size distributions

Once statistically significant Pareto rank-size distributions were recorded, it became necessary to test the time dependence of such distributions. The database of Nel and Hill[6] was used to extract population and enterprise data for the 12 EC Karoo towns and their associated rural areas for selected years during 1911 to 2004. Twelve of the years were selected for the population analyses and ten years were selected for the enterprise analyses.

Are regions unique?

Once statistically significant Pareto rank-size distributions were recorded, it also became necessary to test the uniqueness of a region's population and enterprise distribution. It is possible that the towns in a region are analogous to a random selection of towns from a database. To test this possibility, six random selections were made of towns from a database of 206 South African towns that contains, among others, 2011 population and recent enterprise data for each town. The selected towns and their population and enterprise numbers are presented in Table 2. The rank-size distributions of population and enterprise numbers of each selection were analysed as described earlier.

It was hypothesised that if there are links between the towns of a region resulting in unique rank-size distributions, the same would not be observed for random selections of towns. The null hypothesis was that a random selection of towns from the database would not result in statistically significant rank-size distributions and/or dissimilar regression coefficients.

Results

Population rank-size distributions

Pareto power laws describe the rank-size relationships of populations of the towns of the three regions for the respective years (Table 3). Figure 1 illustrates this relationship for the towns of the GCBR in 2013/2014. In all cases, except the GCBR towns in 1946, more than 90% of the variation was explained (see R^2 in Table 3). There is clearly significant orderliness in the population agglomeration patterns of the South African regions investigated.

Table 2: Six groups of 15 towns each randomly selected to examine their rank-size distributions

No.	Town	Population	Enterprises	Town	Population	Enterprises	Town	Population	Enterprises
	Group 1			**Group 3**			**Group 5**		
1	Alexander Bay	1736	55	Alexander Bay	1736	55	Boshof	8509	36
2	Augrabies	3627	41	Britstown	5145	27	Bredasdorp	15 524	274
3	Botshabelo	181 712	203	Carnarvon	6612	78	De Aar	29 990	223
4	Brandvlei	2859	22	Gariepdam	1568	22	Edenburg	6460	26
5	Christiana	20 882	137	Hennenman	24 355	120	Fraserburg	3029	35
6	Groblershoop	4938	51	Hertzogville	9423	26	Garies	2105	26
7	Hartswater	10 465	295	Keimouth	291	30	Hertzogville	9423	26
8	Hendrina	15 871	85	Middelburg (EC)	18 861	174	Jagersfontein	5729	20
9	Hofmeyr	3680	21	Phuthaditjhaba	54 661	409	Kroonstad	97 780	701
10	Jansenville	5612	75	Richards Bay	252 968	2126	Memel	7142	32
11	Ladismith	7127	108	Springfontein	3699	20	Middelburg (EC)	18 861	174
12	Ladybrand	25 816	258	Sutherland	2836	52	Norvalspont	1198	8
13	Norvalspont	1198	8	Vanderkloof	1228	18	Paul Roux	6152	17
14	Paul Roux	6152	17	Victoria West	8254	88	Springfontein	3699	20
15	Wepener	9553	37	Winterton	6030	117	Steynsburg	7212	42
	Group 2			**Group 4**			**Group 6**		
1	Beaufort-West	71 011	489	De Rust	3566	54	Bethlehem	76 667	993
2	Carnarvon	6612	78	Fauresmith	3628	20	Daniëlskuil	13 597	85
3	Edenburg	6460	26	Kleinmond	6634	210	Douglas	20 083	127
4	Fort Beaufort	25 668	108	Koffiefontein	10 402	39	Gansbaai	11 598	254
5	Greyton	2780	59	Lady Frere	4024	35	Gariepdam	1568	22
6	Keimouth	291	30	Lime Acres	4408	42	Hotazel	1756	16
7	Koffiefontein	10 402	39	Loeriesfontein	2744	29	Koppies	13 803	68
8	Lime Acres	4408	42	Memel	7142	32	Memel	7142	32
9	Phalaborwa	109 468	543	Mossel Bay	89 430	1949	Nieu-Bethesda	1540	58
10	Prieska	14 246	108	Odendaalsrus	63 743	189	Nieuwoudtville	2093	30
11	Reitz	20 183	133	Parys	45 746	506	Sannieshof	11 016	84
12	Sutherland	2836	52	Reddersburg	4886	26	Thabazimbi	28 847	323
13	Swellendam	17 537	398	Schweizer-Reneke	41 226	224	Viljoenskroon	31 468	143
14	Vredendal	18 170	351	Somerset East	18 825	200	Vosburg	1259	16
15	Wepener	9553	37	Viljoensdrif	751	18	Williston	3368	32

The exponent of Zipf's law for the population distribution of cities is normally -1 or close to it.[23] The regions investigated here do not rigidly exhibit Zipf's law because their coefficients are lower than -1 (Table 3). Their Pareto rank-size distributions are nevertheless reasonably close to Zipf's law and predict that a lower-ranked town in the three regions investigated here would have from 55% to 62% of the population of a town just above it in the rank (Table 3).

To examine the time dependence of the population rank-size distribution, use was made of the database for the 12 EC Karoo towns of Nel and Hill.[6] Urban and rural population data recorded by Nel and Hill[6] of 12 different years between 1911 and 2004 for the 12 towns were subjected to the same rank-size analyses described earlier. The results are presented in Table 4.

With the exception of 1911, more than 90% of the variation for every year was explained (see R^2 in Table 4). The results substantiate the finding that the population rank-size distribution follows a Pareto power law (Table 4) and indicate that the relationship holds true over time. Changes in the regression coefficient show a definite pattern. It changed from -1.2 in the early 1900s to about -0.9 by the early 2000s. As a consequence, the population ratio of a lower-ranked town to a town ranked just above it has changed from about 40% in the early 1900s to just over 50% by the early 2000s (Table 4).

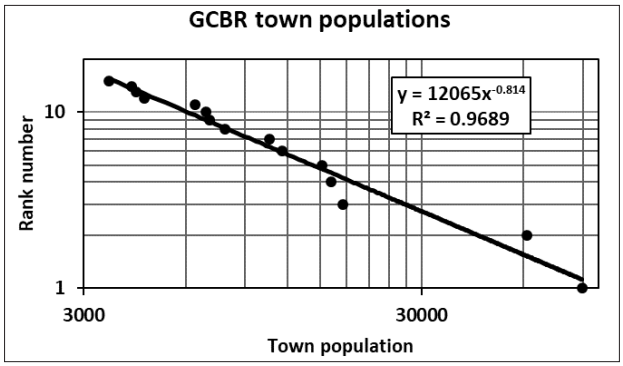


Figure 1: Example of a power law describing the rank-size relationship of population numbers (in 2011) of towns in the Gouritz Cluster Biosphere Reserve (GCBR), South Africa. The GCBR is used as an example.

Table 3: Population rank-size distributions of three South African regions for 1946 and 2011. Towns were ranked according to population sizes, and then regressed as $\log_{10}$ values against $\log_{10}$ values of rank numbers.

Region	Time period	Correlation	R^2	Regression coefficient	*n*	Ratio (%)†
Eastern Cape Karoo	1946	-0.97*	0.943	-0.87	12	54.7
	2011	-0.96*	0.914	-0.86	12	55.1
Strategic environmental assessment study area	1946	-0.97*	0.943	-0.86	28	55.1
	2011	-0.96*	0.919	-0.75	29	59.1
Gouritz Cluster Biosphere Reserve	1946	-0.90*	0.808	-0.70	15	61.6
	2011	-0.97*	0.944	-0.75	15	59.1

**Statistically significant at* $p<0.01$.

†The percentage ratio of the population of a lower ranked town to the population of the town ranked just above it.

Table 4: Population rank-size distribution analyses spanning the period 1911 to 2004 of the 12 Eastern Cape Karoo towns. Towns were ranked according to population sizes of their urban and rural areas, which were regressed as $\log_{10}$ values against $\log_{10}$ values of rank numbers.

Year	Correlation	R^2	Regression coefficient	*n*	Ratio (%)†
1911	-0.91*	0.836	-1.19	12	44
1921	-0.95*	0.902	-1.27	12	42
1936	-0.96*	0.915	-1.19	12	44
1946	-0.96*	0.918	-1.11	12	46
1951	-0.96*	0.917	-1.07	12	48
1960	-0.96*	0.926	-1.07	12	48
1970	-0.96*	0.924	-1.09	12	47
1980	-0.97*	0.942	-1.03	12	49
1985	-0.96*	0.922	-0.99	12	50
1991	-0.95*	0.906	-0.93	12	50
2001	-0.96*	0.915	-0.9	12	54
2004	-0.96*	0.915	-0.89	12	54

**Statistically significant at* $p<0.01$.

†The percentage ratio of the population of a lower ranked town to the population of the town ranked just above it.

Enterprise number rank-size distributions

The rank-size relationships of enterprise numbers in 1946/1947 and 2013/2014 of the towns of the three regions and the ranks of their enterprise numbers are also described by power laws (Table 5). Figure 2 shows the relationship for the GCBR towns in 2013/2014. Except for the GCBR towns in 1946/1947, more than 90% of the variation was explained for both time periods (see R^2 in Table 5). This result is not unexpected given the often-observed linear relationship between population sizes and enterprise numbers of South African towns.[7-19] If the rank-size population distribution of towns in a region is described by a power law, the enterprise number rank-size distribution of the region should exhibit the same pattern.

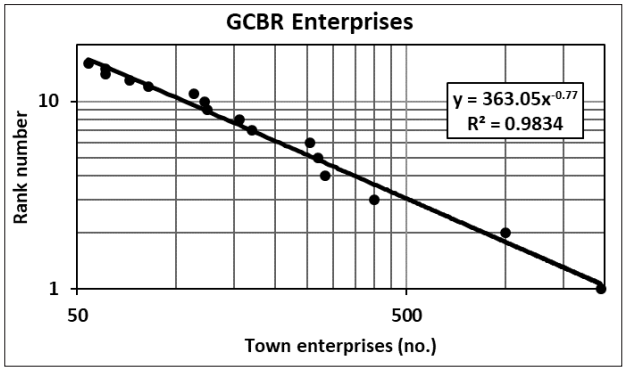


Figure 2: Example of a power law describing the rank-size relationship of enterprise numbers (in 2013/2014) of towns in the Gouritz Cluster Biosphere Reserve (GCBR), South Africa. The GCBR is used as an example.

Table 5: Enterprise number rank-size distributions of three South African regions for 1946 and 2011. Towns were ranked according to their enterprise numbers, and regressed as $\log_{10}$ values against $\log_{10}$ values of rank numbers.

Region	Time period	Correlation	R^2	Regression coefficient	*n*	Ratio (%)†
Eastern Cape Karoo	1946/1947	-0.96*	0.924	-0.98	12	51
	2015/2016	-0.96*	0.918	-0.68	12	63
Strategic environmental assessment study area	1946/1947	-0.95*	0.907	-1.12	28	46
	2015/2016	-0.98*	0.951	-0.76	29	59
Gouritz Cluster Biosphere Reserve	1946/1947	-0.92*	0.847	-0.88	15	54
	2013/2014	-0.99*	0.980	-0.77	15	59

**Statistically significant at* $p<0.01$.

†The percentage ratio of the population of a lower ranked town to the population of the town ranked just above it.

The ratios observed for the enterprise distribution of the three regions indicate that lower-ranked towns in the three regions have from 50% to over 60% of the enterprises of the towns ranked just above them (Table 5). The ratios of all towns in 2015/2016 were higher than those of 1946/1947, suggesting a shift over time. Based on these ratios there is no clear distinction between the different regions (Table 3).

To examine the time dependence of the enterprise number Pareto rank-size distributions of the EC Karoo towns, use was also made of the database for the 12 EC Karoo towns of Nel and Hill[6]. Enterprise numbers recorded by Nel and Hill[6] of 10 years between 1904 and 2000 for the

12 towns were subjected to the same rank-size analyses described earlier. The results are presented in Table 6.

Pareto-like rank-size distributions of enterprises in the EC Karoo towns were observed between 1904 and 2000. Pareto rank-size distributions appear to be enduring and time-independent characteristics. The ratios observed for the enterprise distributions indicate that lower-ranked towns in the region varied from 42% to over 64% of the enterprises of the towns ranked just above them (Table 6). A definite shift in the ratio was observed after 1951: lower-ranked towns progressively had more enterprises relative to higher-ranked towns. This shift could be because lower-ranked towns had over time increased populations relative to higher-ranked towns (Table 3).

Table 6: Enterprise number rank-size distributions of Eastern Cape Karoo towns for the period 1904 to 2000. Towns were ranked according to enterprise numbers, which were regressed as $\log_{10}$ values against $\log_{10}$ values of rank numbers.

Year	Correlation	R^2	Regression coefficient	*n*	Ratio (%)†
1904	-0.96*	0.927	-1.06	12	48
1911	-0.96*	0.926	-1.26	12	42
1921	-0.95*	0.907	-1.22	12	43
1935	-0.95*	0.900	-1.05	12	48
1951	-0.97*	0.935	-1.16	12	45
1961	-0.97*	0.948	-0.91	12	53
1970	-0.96*	0.923	-0.86	12	55
1980	-0.97*	0.941	-0.85	12	55
1990	-0.95*	0.912	-0.74	12	60
2000	-0.95*	0.905	-0.65	12	64

**Statistically significant at* $p<0.01$.

†The percentage ratio of the population of a lower ranked town to the population of the town ranked just above it.

Distributions of randomly selected groups of towns

The rank-size population distributions of six randomly selected groups of South African towns exhibited Pareto-like properties and the correlations are statistically significant (Table 7). The regression coefficients varied from -0.48 to -0.70 with an average of -0.58 ± 0.073, which is clearly lower than the regression coefficients of the three selected regions (Table 3). The average is also much lower than the century-long coefficients of the towns of the EC Karoo (Table 4). The ratios of more than 60% of lower-ranked towns to towns ranked just above them are consequently higher than those of the three regions (compare Table 7 with Tables 3 and 4). In randomly selected groups of towns, population numbers tend to be more evenly distributed.

The rank-size enterprise distributions of the selected groups also exhibited Pareto-like properties and the correlations are statistically significant (Table 7). The regression coefficients of the randomly selected groups varied from -0.56 to -0.71 with an average of -0.64 ± 0.0059, which is somewhat lower than the regression coefficients of the three selected regions (Table 5). These coefficients are also lower than the century-long coefficients of the EC Karoo towns except for the year 2000. As a consequence, the ratios of lower-ranked towns to towns ranked just above them are higher (compare Tables 5 and 7). The enterprise numbers of randomly selected groups of towns were more evenly distributed than those of the three regions.

Because the database of more than 200 towns contains large as well as small towns, the random selection of 15 towns from it should yield a group that has a spread of town sizes (measured in terms of populations or enterprise numbers). This spread could lead to the recording of statistically significant log-log rank-size distributions, as was observed (Table 7). The coefficients of the randomly selected groups were, however, lower and their ratios higher than those of the three regions, which suggests that population agglomeration and enterprise development patterns of regions do have uniqueness. Expressed differently: what happens in one town of a specific region influences what happens in other towns of that region. Regions, however, do not appear to differ from one another.

Returning to the questions initially raised in this contribution, the following can be concluded:

- Statistically significant log-log rank-size distributions apply to population agglomerations of towns in different regions of South Africa.
- Such rank-size distributions also apply to the number of enterprises in these towns.
- Based on the relationships observed, there is no clear distinction between different regions. However, the regional agglomerations differed from those of groups of towns randomly selected from a database (compare Tables 5 and 7). Regions, therefore, appear to have some uniqueness regarding such agglomerations.

Table 7: The rank-size distributions of populations and enterprise numbers of six groups of 15 towns each randomly selected from a large database

Group	Populations				Enterprises			
	Correlation	R^2	Regression coefficient	Ratio (%)†	Correlation	R^2	Regression coefficient	Ratio (%)†
Group 1	-0.98*	0.951	-0.62	65	-0.93*	0.861	-0.69	62
Group 2	-0.88*	0.776	-0.48	72	-0.96*	0.918	-0.71	61
Group 3	-0.96*	0.979	-0.59	66	-0.99*	0.979	-0.59	66
Group 4	-0.95*	0.896	-0.55	68	-0.98*	0.959	-0.56	68
Group 5	-0.97*	0.931	-0.70	62	-0.97*	0.939	-0.61	66
Group 6	-0.94*	0.877	-0.56	68	-0.98*	0.970	-0.65	64

**Statistically significant at* $p<0.01$.

†The percentage ratio of the population of a lower ranked town to the population of the town ranked just above it.

Discussion

Population distribution across geographical areas is not random: there is a strong tendency toward agglomeration.[1] Why then are there large and small towns? This question led Christaller to theorise about the centrality of towns, based on the services that towns deliver to their hinterlands.[33] Christaller argued about a system of central places that exhibits a hierarchical principle: any goods supplied in a central place of order *i* is also supplied in all central places of order $j>i$. Centrality became an important issue in studies of South African towns.[34]

However, Eaton and Lipsey[35] argued that Christaller's theory of central places is simultaneously a theory of the location and agglomeration of economic activity in which there is no force creating agglomeration, in which agglomeration serves no purpose, and in which no firm ever chooses a location. Despite this criticism, economic geographers and regional economists remain interested in the reasons for uneven regional development. They still ask why economic growth does not lead to similar levels of prosperity, employment and welfare across space (for example see Gardiner et al.[36]); a question of relevance also about South African regions.

A number of questions were posed in this study that can now be answered. Whereas population agglomeration patterns in many countries are subject to Zipf's law, a power law with Pareto-like characteristics[1,2], this study has demonstrated that Pareto-like population and enterprise distributions close to Zipf's law are present in the towns of different South African regions (Tables 3 and 5) – a fact hitherto unknown. Morudu and du Plessis[5] reported Zipf-like distribution of the population, employment and economic value addition characteristics of South African municipalities. The latter are organisations that are man-made constructs, often containing more than one town, which do not necessarily reflect the 'natural' way in which agglomeration phenomena evolve in towns of a region.

The Pareto distribution types encountered in this study have endured over a long time even when the regression coefficients progressively changed (Tables 3 and 6). The patterns recorded in regions differ from that of randomly selected groups of towns, suggesting that there is some uniqueness in the orderly way in which regions give rise to population and enterprise agglomeration patterns.

This study has revealed an additional dimension of the regularities observed between population and/or enterprise characteristics in South African towns.[7-19] Previously, linear relationships between population and enterprise numbers and enduring power law relationships between total enterprise numbers and the enterprise richness (total number of enterprise types) of South African towns[15,19] were recorded. This study recorded log-log agglomeration patterns for the population and enterprise numbers of regional towns. Population growth and distribution, and enterprise development and distribution, are clearly highly orderly processes. This fact should be factored into local economic development planning and support, as was also suggested by Morudu and du Plessis[5].

The similarities in the variance patterns for population and enterprise distributions observed in this study, raise a 'chicken or egg' scenario, that is, does population growth precede enterprise development, or does enterprise development precede population growth? Fujita and Thisse[37] argue that consumer behaviour predicts agglomeration because consumers face search costs and have incomplete information about the retail landscape, so they find it efficient to patronise larger centres. Firms cluster because of consumer behaviour and benefit from demand externalities by locating in the larger centres. Their thinking, therefore, implies that population growth precedes enterprise development. Eaton and Lipsey[35] also commented that because the clustering of heterogeneous firms facilitates multipurpose shopping, it allows consumers to economise on shopping costs.

However, the contrary – that enterprise development could precede population growth – should also be considered. Gabaix[2] stated that the creation of jobs is important and people are attracted to where there are jobs. This implies that enterprise growth has to take place before population growth results from immigration of people seeking employment. Eeckhout[1] remarked that agglomeration and residential mobility of the population between different geographical locations are tightly connected to economic activity and that the evolution of populations across geographical locations is an extremely complex amalgam of incentives and actions taken by many individuals, enterprises and organisations.

Fransen[29] remarked that most towns in the former Cape Colony in South Africa were not founded for commercial reasons. So-called 'church towns' developed around churches that were built to satisfy the needs of farming communities for religious services. In many cases following the building of a church, members of a congregation built 'town houses' for use when the rural families attended church services. Regular gathering of people in these settings created informal markets where goods were exchanged or bartered. These markets attracted entrepreneurs and the establishment of enterprises followed. In this case, population growth preceded enterprise growth. However, over time, the presence of enterprises and the possibility of finding employment, attracted more people to the fledgling towns, thereby growing the local market.

Population growth and enterprise development, therefore, seem to proceed hand in hand. On the one side, an entrepreneur might start a new business and if it is successful, it contributes to employment, which enhances the image of the town as a place to find employment. On the other side, a new immigrant attracted to the town enhances its total disposable income, thereby increasing the ability of the town to 'carry' more enterprises. This interdependence probably causes the similarities of the population and enterprise distribution patterns. This fits in with Eeckhout's suggestion that agglomeration and residential mobility of the population between different geographical locations are tightly connected to economic activity.[1]

However, a better understanding of the links between population and enterprise growth in regional towns still does not explain fully why some towns grow large and others stay small. Krugman's[3] lament that there is not a good theory to account for the striking empirical regularities in population agglomeration patterns observed, which is now also the case for enterprise agglomeration patterns, still applies. There is a need to further explore these issues. The extreme skewness in population and enterprise numbers of different towns' distributions should, however, be considered in local economic development planning and execution.

Acknowledgements

Funding and support for this study was received from the Board of the Gouritz Cluster Biosphere Reserve and the Centre for Environmental Management, University of the Free State. Etienne Nel, Trevor Hill and Doreen Atkinson allowed access to their database on the towns of the Eastern Cape Karoo.

References

1. Eeckhout J. Gibrat's law for (all) cities. Am Econ Rev. 2004;94(5):1429–1451. https://doi.org/10.1257/0002828043052303
2. Gabaix X. Zipf's law for cities: An explanation. Q J Econ. 1999;114(3):739–767. https://doi.org/10.1162/003355399556133
3. Krugman P. Confronting the mystery of urban hierarchy. J Japan Int Econ. 1996;10(4):399–418. https://doi.org/10.1006/jjie.1996.0023
4. Andriani P, McKelvey B. From Gaussian to Paretian thinking: Causes and implications of power laws in organizations. Organ Sci. 2009;20(6):1053–1071. https://doi.org/10.1287/orsc.1090.0481
5. Morudu H, Du Plessis D. Economic and demographic trends of municipalities in South Africa: An application of Zipf's rule. Town and Regional Planning. 2013;63:24–36. http://dx.doi.org/10.13140/2.1.2587.7125
6. Nel E, Hill T. Marginalisation and demographic change in the semi-arid Karoo, South Africa. J Arid Env. 2008;72:2264–2274. https://doi.org/10.1016/j.jaridenv.2008.07.015
7. Toerien DF, Seaman MT. The enterprise ecology of towns in the Karoo, South Africa. S Afr J Sci. 2010;106(5/6):24–33. http://dx.doi.org/10.4102/sajs.v106i5/6.182

8. Toerien DF, Seaman MT. Proportionality in enterprise development of South African towns. S Afr J Sci. 2012;108(5/6):38–47. http://dx.doi.org/10.4102/sajs.v108i5/6.588

9. Toerien DF, Seaman MT. Regional order in the enterprise structures of selected EC Karoo towns. S Afr Geogr J. 2012;94(2):1–15. http://dx.doi.org/10.1080/03736245.2012.742782

10. Toerien DF, Seaman MT. Evidence of island effects in South African enterprise ecosystems. In: Mahamane A, editor. The functioning of ecosystems. Rijeka: Intech; 2012. p. 229–248. http://dx.doi.org/10.5772/36641

11. Toerien DF. 'n Eeu van orde in sakeondernemings in dorpe van die Oos-Kaapse Karoo [A century of order in the enterprises of towns of the Eastern Cape Karoo]. LitNet Akad. 2014;11(1):330–371. Afrikaans.

12. Toerien DF. The enterprise architecture of Free State towns. Technical report 2014 DTK [document on the Internet]. c2014 [cited 2017 Jun 28]. Available from: https://www.researchgate.net/profile/Daan_Toerien/publications

13. Toerien DF, Seaman MT. Ecology, water and enterprise development in selected rural South African towns. Water SA. 2011;37(1):47–56. https://doi.org/10.4314/wsa.v37i1.64106

14. Toerien DF, Seaman MT. Paradoxes, the tyranny of structures and enterprise development in South African towns. Paper presented at: Strategies to overcome poverty and inequality: Towards Carnegie3; 2012 September 3–7; Cape Town, South Africa. Available from: http://carnegie3.org.za/docs/papers/269_Toerien_Paradoxes,%20the%20tyranny%20of%20structures%20and%20enterprise%20development%20in%20SA%20towns.pdf

15. Toerien DF, Seaman MT. Enterprise richness as an important characteristic of South African towns. S Afr J Sci. 2014;110(11/12), Art. #2014-0018, 9 pages. http://dx.doi.org/10.1590/sajs.2014/20140018

16. Toerien DF. Enterprise proportionalities in the tourism sector of South African towns. In: Kasimoglu M, editor. Visions of global tourism industry: Creating and sustaining competitive strategies. Rijeka: Intech; 2012. p. 113–138. http://dx.doi.org/10.5772/37319

17. Toerien DF. Economic value addition, employment, and enterprise profiles of local authorities in the Free State, South Africa. Cogent Soc Sc. 2015;1, Art. #1054610, 21 pages http://dx.doi.org/10.1080/23311886.2015.1054610

18. Toerien DF. New utilization/conservation dilemmas in the Karoo, South Africa: Potential economic, demographic and entrepreneurial consequences. In: Ferguson G, editor. Arid and semi-arid environments: biogeodiversity, impacts and environmental challenges. New York: Nova Science Publishers; 2015. p. 79–123.

19. Toerien DF. The enduring and spatial nature of the enterprise richness of South African towns. S Afr J Sci. 2017;113(3/4), Art. #2016-0190, 8 pages. https://doi.org/10.17159/sajs.2017/20160190

20. Knudsen T. Zipf's law for cities and beyond: The case of Denmark. Am J Econ Sociol. 2001;60(1):124–146. https://doi.org/10.1111/1536-7150.00057

21. Scholes R, Lochner P, Schreiner G, Snyman-Van der Walt L, De Jager M, editors. Shale gas development in the Central Karoo: A scientific assessment of the opportunities and risks. Stellenbosch: Council for Scientific and Industrial Research; 2016. Available from: http://seasgd.csir.co.za/wp-content/uploads/2016/12/SGD-Scientific-Assessment-Binder1_LOW-RES_INCL-ADDENDA_21Nov2016.pdf

22. Toerien D. Enterprise development and dynamics in the Gouritz Cluster Biosphere Reserve. Special Report (undated): Gouritz Cluster Biosphere Reserve [document on the Internet]. No date [cited 2017 Jun 25]. Available from: http://www.gouritz.com/enterprise-development-and-dynamics-ingcbr-%20towns-hot-off-the-press-repor%20t-available-a-quantitative-study-ofenterprise-%20development-and-dyna#.WuwqdliFOUl

23. Brakman S, Garretsen H, Van Marrewijk C, Van den Berg M. The return of Zipf: Towards a further understanding of the rank-size distribution. J Regional Sci. 1999;39(1):183–213. https://doi.org/10.1111/1467-9787.00129

24. Blank A, Solomon S. Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components). Physica A. 2000;287:279–288. https://doi.org/10.1016/S0378-4371(00)00464-7

25. Axtell RL. Zipf distribution of U.S. firm sizes. Science. 2001;293(5536):1818–1820. https://doi.org/10.1126/science.1062081

26. Florence PS. The logic of British and American industry: A realistic analysis of economic structure and government. Chapel Hill, NC: University of North Carolina Press; 1953.

27. Axtell RL. Firm sizes: Facts, formulae, fables and fantasies. SSRN Electr J. 2006, working paper #44, 23 pages. https://doi.org/10.2139/ssrn.1024813

28. Nel E, Taylor B, Hill T, Atkinson D. Demographic and economic changes in small towns in South Africa's Karoo: Looking from the inside out. Urban Forum. 2011;22:395–410. https://doi.org/10.1007/s12132-011-9131-z

29. Fransen H. Old towns and villages of the Cape. A survey of the origin and development of towns, villages and hamlets at the Cape of Good Hope. Johannesburg: Jonathan Ball Publishers; 2006.

30. Burman J. The Little Karoo. Cape Town: Human & Rousseau; 1981.

31. Population of South Africa: 1904–1970. Pretoria: Government Printing Works; 1976.

32. City population: South Africa [webpage on the Internet]. No date [cited 2017 May 25]. Available from: https://www.citypopulation.de/php/southafrica

33. Baskin CW. Central places in southern Germany. Translation of Christaller W: Die Zentralen Orte in Süddeutschland. Englewood: Prentice-Hall; 1966.

34. Van der Merwe IJ, Nel E. Die stad en sy omgewing: 'n studie in nedersettingsgeografie [The city and its environment: A study of settlement geography]. Stellenbosch: Universiteits-uitgewers; 1975. Afrikaans.

35. Eaton C, Lipsey RG. An economic theory of central places. Econ J. 1982;92(365):56–72. https://doi.org/10.2307/2232256

36. Gardiner B, Martin R, Tyler P. Does spatial agglomeration increase national growth? Some evidence from Europe. J Econ Geogr. 2011;11(6):979–1006. https://doi.org/10.1093/jeg/lbq047

37. Fujita M, Thisse J. Economics of agglomeration: Cities, industrial location, and regional growth. New York: Cambridge University Press; 2002. https://doi.org/10.1017/CBO9780511805660

Arable agriculture changes soil microbial communities in the South African Grassland Biome

AUTHORS:
Gilbert Kamgan Nkuekam[1]
Don A. Cowan[1]
Angel Valverde[1,2]

AFFILIATIONS:
[1]Centre for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, Pretoria, South Africa
[2]Department of Microbial, Biochemical and Food Biotechnology, University of the Free State, Bloemfontein, South Africa

CORRESPONDENCE TO:
Gilbert Kamgan Nkuekam

EMAIL:
kamgan2000@yahoo.com

DATES:
Received: 21 Aug. 2017
Revised: 21 Nov. 2017
Accepted: 16 Mar. 2018
Published: 30 May 2018

KEYWORDS:
land use; grassland; agriculture; microbial diversity; next-generation sequencing

HOW TO CITE:
Kamgan Nkuekam G, Cowan DA, Valverde A. Arable agriculture changes soil microbial communities in the South African Grassland Biome. S Afr J Sci. 2018;114(5/6), Art. #2017-0288, 7 pages. http://dx.doi.org/10.17159/sajs.2018/20170288

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa); University of Pretoria

Many studies, mostly in temperate regions of the northern hemisphere, have demonstrated that agricultural practices affect the composition and diversity of soil microbial communities. However, very little is known about the impact of agriculture on the microbial communities in other regions of the world, most particularly on the African continent. In this study, we used MiSeq amplicon sequencing of bacterial 16S rRNA genes and fungal ITS regions to characterise microbial communities in agricultural and natural grassland soils located in the Mpumalanga Province of South Africa. Nine soil chemical parameters were also measured to evaluate the effects of edaphic factors on microbial community diversity. Bacterial and fungal communities were significantly richer and more diverse in natural grassland than in agricultural soils. Microbial taxonomic composition was also significantly different between the two habitat types. The phylum Acidobacteria was significantly more abundant in natural grassland than in agricultural soils, while Actinobacteria and the family Nectriaceae showed the opposite pattern. Soil pH and phosphorus significantly influenced bacterial communities, whereas phosphorus and calcium influenced fungal communities. These findings may be interpreted as a negative impact of land-use change on soil microbial diversity and composition.

Significance:

- This report is the first of the effect of land-use changes on the diversity of the soil microbial communities in African grassland soils.
- Land-use changes influence the diversity and structure of soil microbial communities in the Grassland Biome of South Africa.
- This study serves as a baseline for future studies on South African soil microbial diversity.

Introduction

Soils represent a reservoir for a wide diversity of microorganisms such as bacteria, fungi and viruses.[1] Bacteria and bacteriophages are typically the most abundant microorganisms present in soils, although their prevalence is highly variable and affected by edaphic factors such as soil mineral content and pH[1,2] and by local plant biodiversity[3]. Soil microbial communities are important drivers of ecosystem functioning and climate change mitigation through the fixation, immobilisation and cycling of greenhouse gases.[4] Plants synthesise organic matter via photosynthetic activities and provide energy to soil microbes through root exudates.[5,6] In return, soil microorganisms provide the plants with critical 'services', including decomposition of organic matter, mineral cycling and biocontrol of soil-borne pathogens.[4,7]

Land use, such as the modification of a natural ecosystem for agriculture purposes, has been proven to have significant effects on soil microbial communities by changing the physical and chemical properties of the soil. Such effects on the microbial community include changes in microbial species abundance, richness and diversity.[8-11] The effect of changing land use on the diversity of soil microbes has also been observed in grassland ecosystems. For example, Acosta-Martinez et al.[12] studied soil bacteria diversity in a single soil type in Texas, USA. They found much higher bacteria diversity in soils under agriculture than in natural grassland soils, providing evidence of the positive influence of land use on soil bacterial diversity. The overall bacterial community diversity and composition in different grassland soils from across South and North America were significantly influenced with soil pH.[13] Similar results were reported in a study of bacterial communities in German grassland and forest soils.[14,15] In a more recent study, both bacterial and fungal communities exhibited contrasting beta diversity among two types of European subalpine/alpine grasslands, and both bacterial and fungal communities were influenced by grassland type.[16] However, most of these studies have been conducted in temperate northern hemisphere grasslands and none is directly relevant to the unique and defined biomes of South Africa.[17] We argue that more geographically diverse comparative analyses are required in order to better understand how microbial communities are altered by land use.

South Africa is one of the world's biodiversity hotspots, and is ranked third in the global list; it is composed of nine biomes which together contain between 250 000 and 1 000 000 species including animals and plants, many of which are endemic to the country.[17] However, while higher eukaryotes have been intensively studied in southern Africa, there has been only very limited focus on the biodiversity of soil microbial communities in any of South Africa's major biomes. Mpumalanga is a province in the northeast of South Africa, much of which is designated as Grassland Biome.[17] The Grassland Biome is the cornerstone of commercial maize cropping, and many grassland areas have been converted to arable production. Sorghum, wheat and sunflowers are also farmed in the region but on a smaller scale.[18] Agriculture is, therefore, one of the important economic sectors in Mpumalanga, with significant contributions toward national economic growth[19], but with major modifications to much of the original natural grassland habitat.

In this study, we evaluated the differences in bacterial and fungal communities across a series of paired agricultural and natural grassland soils in the Mpumalanga Province, South Africa. We used an amplicon sequencing based approach to characterise the soil bacterial and fungal communities and to quantify their differences across 15 sites, and related these differences to measured soil chemical parameters. Inventories of the composition and differences in the bacterial and fungal communities in Mpumalanga's agricultural and natural grassland soils will contribute to the body of knowledge, and can provide a description of the core soil microbial communities in the unmodified grassland.

Materials and methods

Sample collection

The study sites are situated on the Highveld escarpment in the Mpumalanga Province, which makes up 6.5% of South Africa's land area, and form part of the Grassland Biome.[20] The climatic conditions of the Grassland Biome of South Africa span a rainfall gradient from about 400 mm to >1200 mm per year, a temperature gradient from frost-free to snow-bound in winter, and an altitude from sea level to >3300 m; a spectrum of soil types occurs in the Grassland Biome, from humic clays to poorly structured sands.[21] In April 2016, 15 paired agricultural and natural sites in the Grassland Biome were localised using a geographic information system (GIS) technology (Supplementary table 1). The distance between sampling sites spanned 18–240 km, and at each sampling site, the sampling points between the two habitat types (agricultural vs natural grassland soil) were approximately 200 m apart. At each GPS-located sampling site, five pseudo-replicate soil samples (50–200 mm depth) were recovered from within a 10 x 10 m quadrat. All pseudo-replicate samples from each habitat type (agricultural vs natural grassland soil) were bulked into a single sample. This process was repeated for all 15 sampling sites, resulting in a total of 30 soil samples (15 agricultural and 15 grassland soils). The samples were transported to the laboratory at 5 °C and stored at -20 °C for subsequent DNA extraction.

Analysis of the soil chemical parameters

Nine soil chemical parameters were analysed to evaluate their influence on the structure and composition of the soil bacterial and fungal communities. Soil pH was measured using a Crison pH meter (Crison Instruments, Barcelona, Spain) from the supernatants of 1:2.5 (g/g) soil/deionised water suspensions, after soil particles had settled after 2 min of vigorous shaking.[22] Total inorganic phosphate was measured from the supernatant of 4 g of soil in 39 mL of the Bray-1 solution after 1 min of vigorous shaking and filtration through Whatman paper.[23,24] Extractable cations (K, Ca, Na, Mg) were measured from the supernatants of 1:10 (g/g) soil/ammonium acetate suspensions.[24] Total carbon content was determined by oxidation with potassium dichromate and sulfuric acid as described previously.[24] Total ammonium (NH_4) and nitrate (NO_3) contents were analysed by Bemlab (Pty) Ltd (Strand, Western Cape, South Africa) using standard protocols.

Genomic DNA extraction, amplification and high-throughput sequencing

Genomic DNA was extracted from 0.25 g soil (dry weight) using the PowerSoil DNA Isolation Kit (Mo Bio Laboratories Inc., Carlsbad, CA, USA) within 48 h of sampling, following the manufacturer's instructions. Soil samples were ground with an electric Powerlyser (Mo Bio Laboratories Inc.) before further processing. Aliquots of 5 μL of genomic DNA were premixed with GelRedTM Nucleic Acid Gel Stain (Biotium, Hayward, CA, USA), separated on 1% agarose gels and visualised under UV light to determine the success of the extraction.

DNA amplification was done at MRDNA (www.mrdnalab.com, Shallowater, TX, USA) sequencing facility in a 30-cycle PCR, using the HotStarTaq Plus Master Mix Kit (Qiagen, Germantown, MD, USA). For bacteria, 16S rRNA gene V4 variable region PCR primers 515/806[25] were used. For fungi, internal transcribed spacer region (ITS) primers ITS1f and ITS4[26] were used. The thermal cycling conditions for amplification of the two gene regions were as follows: 94 °C for 3 min, followed by 28 cycles of 94 °C for 30 s, 53 °C for 40 s and 72 °C for 1 min, with a final elongation step at 72 °C for 5 min. After amplification, PCR products were checked on 2% agarose gels to determine the success of amplification and the relative intensity of bands. Multiple samples were pooled together (e.g. 30 samples) in equal proportions based on their molecular weight and DNA concentrations. Pooled samples were purified using calibrated Ampure XP beads. Pooled and purified PCR products were used to prepare DNA libraries following the Illumina TruSeq DNA library preparation protocol. High throughput sequencing was performed at MRDNA on a MiSeq platform following the manufacturer's guidelines.

Sequence analysis

Sequences were processed using Qiime V1.9.1.[27] Bacterial and fungal sequences were analysed independently. Firstly, the mapping file was checked to ensure that it is formatted correctly. Barcodes were extracted from the sequence reads. The split libraries command was used to demultiplex the fastq file and assign sequence reads to their respective sample according to their corresponding barcodes, using sample mapping information (Supplementary tables 2 and 3) at the default parameter implemented in Qiime. Chimeric sequences were screened and then removed using the USEARCH software[28] and the Greengenes-derived 16S reference database implemented in Qiime[29] for archaea and bacteria, respectively, and the UNITE-INSD (release 7) derived ITS reference database for fungi. High-quality reads were clustered into operational taxonomic units (OTUs) at 97% sequence similarity and assigned taxonomy using blast search against the Greengenes database (for archaea and bacteria) and the UNITE-INSD database (for fungi), respectively. Sequences identified as singletons were removed before sub-sampling to a depth of 3001 (for bacteria) and 12 358 (for fungi) sequences per sample to ensure equal sampling effort across samples.

Data access

The DNA sequences from this study were deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) database, with accession numbers SRR5341506 and SRR5341505 for bacterial 16S and fungal ITS, respectively.

Statistical analyses

All statistical analyses were conducted in R version 3.2.1.[30] Differences in community composition were visualised using non-metric multidimensional scaling with Hellinger-transformed Bray–Curtis dissimilarities. Abiotic data were standardised (mean=0, SD=1) and the resulting data matrices were used to perform principal component analysis to evaluate differences in soil chemistry between agricultural soils and natural grassland soils. An analysis of similarity (ANOSIM) was performed to determine whether differences between habitats (agricultural vs natural grassland soils) were statistically significant (999 permutations, $\alpha<0.05$). The number of shared OTUs between communities or samples was visualised using Venn diagrams. Differences in means for bacterial and fungal diversities metrics (species richness, Shannon–Wiener index, Simpson index and Pielou's evenness), phyla abundance and abiotic data were compared using paired two-tailed Student's *t*-tests. The compositions of major bacterial and fungal phyla were visualised using a heatmap with unweighted pair group method with arithmetic mean (UPGMA) clustering on Bray–Curtis distances after Hellinger transformation. Distance-based redundancy analysis was used to evaluate the effects of the environment on microbial community composition.[31]

Results and discussion

We evaluated and compared the compositions of the bacterial and fungal communities in agricultural and natural grassland soils from 15 sites in the Mpumalanga Province in the Grassland Biome of South Africa using deep amplicon sequencing. We also evaluated the effect of edaphic factors on the patterns of the bacterial and fungal communities in the two habitats.

After quality filtering, removal of chimeras and singletons and sub-sampling, 90 030 bacterial and 370 740 fungal sequences remained, clustering into 42 866 bacterial and 9730 fungal OTUs (Figure 1, Table 1), respectively. Although the relative numbers of bacterial versus fungal taxa cannot be

quantified from our data, our results suggest that natural grassland soils are richer in bacterial taxa than in fungal taxa, which is in agreement with previous studies showing that soil microbial biomass in natural grasslands is usually strongly dominated by bacteria.[32,33]

Bacterial alpha diversity metrics (Shannon and Simpson diversity indices and Pielou's evenness) revealed no significant difference (paired *t*-test, $p>0.05$) between agricultural and natural grassland soils (Table 1). However, natural grassland soils were significantly richer (paired *t*-test, $p<0.05$) in bacterial OTUs than were agricultural soils (Table 1), confirming the effects of land-use perturbation in decreasing soil microbial diversity.[34,35] However, this finding contradicts previous studies on grassland soils, in which higher bacterial diversity was associated with land-use changes.[11,36,37] The high level of bacterial genotypic richness detected in natural grassland may be explained by the influence of the high diversity of plant species in the grassland ecosystem, compared to the monoculture of agricultural crops.[8,38] Plant species differ in the quantity and variety of their root exudates, as well as in the range of organic compounds in plant litter detritus: both factors are potential drivers of the higher microbial diversity often associated with more diverse plant communities.[34,39-41] Our results are therefore consistent with the view that plant diversity promotes soil microbial diversity in grassland ecosystems.[3]

Fungal alpha diversity metrics (species richness, Shannon and Simpson diversity indices, and Pielou's evenness) were significantly higher (paired *t*-test, $p<0.05$) in natural grassland soils than in agricultural soils (Figure 1, Table 1), which is also consistent with previous studies on grasslands.[11,36,37] The relatively low fungal diversity detected in agricultural soils is probably as a result of land-use management processes such as tillage. Tillage has been shown to damage fungal mycelia, leading to a reduction in fungal abundance and diversity in the soil.[42,43] However, these results should be treated with caution as rarefaction curves obtained for both bacteria and fungi (Supplementary figures 1 and 2) did not plateau, indicating that the diversity of bacteria and fungi in both agricultural and natural grassland soils had not been fully sampled.

Differences in microbial diversity in managed agricultural soils, as compared to natural grassland soils, have been disputed in several studies. Some studies reported high bacterial diversity in conventionally managed arable soils as well as grassland soils.[32,33,36,37] Others have shown that extensively managed grassland soils are richer in fungi than bacteria.[37,44]

Figure 1: Venn diagrams of shared and unique (a) bacterial and (b) fungal operational taxonomic units between agricultural and natural grassland soil communities.

Table 1: Characteristics of sequences, diversity and richness between agricultural soils and natural grassland soils

	Bacteria		Fungi	
	Agriculture	**Grassland**	**Agriculture**	**Grassland**
Raw sequences	1 910 115	1 537 765	674 574	502 509
Post pre-processing	80 112	82 105	605 544	453 965
Post sub-sample and filter†	45 015	45 015	185 370	185 370
Number of operational taxonomic units	25 070	24 322	5102	6239
Unclassified	5663 (22.6%)	5015 (20.6%)	104 (2.04%)	116 (1.9%)
Richness	$2116^{a} \pm 242.1$	$2250^{b} \pm 105.4$	$556.4^{a} \pm 176.3$	$748.5^{b} \pm 137.8$
Shannon index	$7.38^{a} \pm 0.37$	$7.54^{a} \pm 0.11$	$3.48^{a} \pm 0.8$	$4.09^{b} \pm 0.6$
Simpson index	$1.00^{a} \pm 0.005$	$1.00^{a} \pm 0.0004$	$0.88^{a} \pm 0.12$	$0.93^{a} \pm 0.006$
Pielou's evenness	$0.96^{a} \pm 0.04$	$0.98^{a} \pm 0.009$	$0.55^{a} \pm 0.11$	$0.62^{a} \pm 0.009$

†Sub-sample depth was at 3001 (for bacteria) and 12 358 (for fungi) sequences per sample. Values are the mean ± standard deviation.

Diversity indices followed by different superscript letters are significantly different (agricultural vs grassland soils) according to paired Student's t-tests ($p<0.05$).

The contrasting pattern of bacterial and fungal diversity in agricultural compared to natural grassland soils has been explained by the difference in nutrient availability in the two ecosystems.[45,46] In extensively managed agricultural soils, high fertility and nutrient availability derived from the input of fertiliser favour bacterial communities, while the low soil fertility characteristics of natural grassland soils favour the fungal communities.[45,46] In comparison, the factors that shape the differences in bacterial and fungal diversity in managed soils are still largely unknown. The bacterial and fungal taxa unique to agricultural soils and natural grassland soils accounted for the majority of the raw sequences (Figure 1, Table 1). In addition, more than 20% of the bacterial sequences and about 2% of the fungal sequences could not be classified (Table 1), suggesting that these soils remain understudied.

A total of 13 prokaryotic phyla (12 bacterial phyla and 1 archaeal phylum) were detected in the two habitats (Figure 2), of which Actinobacteria, Proteobacteria and Acidobacteria were the most abundant (Supplementary table 4). These three bacterial phyla are the most dominant taxa found in soils worldwide and make up the large proportion of bacterial 16S rRNA sequences available in public databases.[47,48] Seven fungal phyla were detected in the two habitats (Figure 3), of which Ascomycota accounted for the majority of the sequences (71.6% and 63.4%) in agricultural and natural grassland soil samples, respectively (Supplementary table 4). Both the bacterial and fungal phyla identified are typically ubiquitous in soils and are thought to be important role players in soil geochemical cycling processes.[3,14]

Species of the phylum Acidobacteria were significantly more abundant (paired *t*-test, $p<0.05$) in natural grassland soils than in agricultural soils (Figure 2, Supplementary table 4). Acidobacteria abundance is commonly reduced in soils under conventional tillage[11,49], which may explain this pattern. This phylum includes many environmentally important species, which provide a wide range of functions including the biodegradation of cellulose, hemicelluloses and chitin, nitrate and nitrite reduction, and the production of antimicrobial compounds.[50]

The most abundant bacterial and fungal taxa accounted for at least 10% of the total number of sequences in both agricultural and natural grassland soil samples. Analysis of these sequences at class, order, family and genus levels is shown in Supplementary table 4. At the class and order levels, Actinobacteria were significantly more abundant (paired *t*-test, $p<0.05$) in agricultural soils than in natural grassland soils (Supplementary table 4). These results suggest that land-use changes favour Actinobacteria abundance, which contradicts the results of previous related studies on non-African grassland soils, which suggested that Actinobacteria are more abundant in non-disturbed grassland soils than in agricultural managed soils.[11,12] Actinobacteria are major producers of extracellular enzymes and secondary metabolites, and are thought to play significant roles in carbon cycling, plant disease suppression and enhancement of plant growth.[51,52]

At family level, Nectriaceae (Hypocreales) was the most abundant fungal family with 17.4% and 8.7% of all the fungal raw sequences in agricultural soils and natural grassland soils, respectively (Supplementary table 4). However, species of the family were significantly more abundant (paired *t*-test, $p<0.05$) in agricultural soils than in natural grassland soils (Supplementary table 4). The family Nectriaceae includes species of both economic and health importance. Species of the genus *Fusarium* produce mycotoxin and are both crop and opportunistic human pathogens.[53,54]

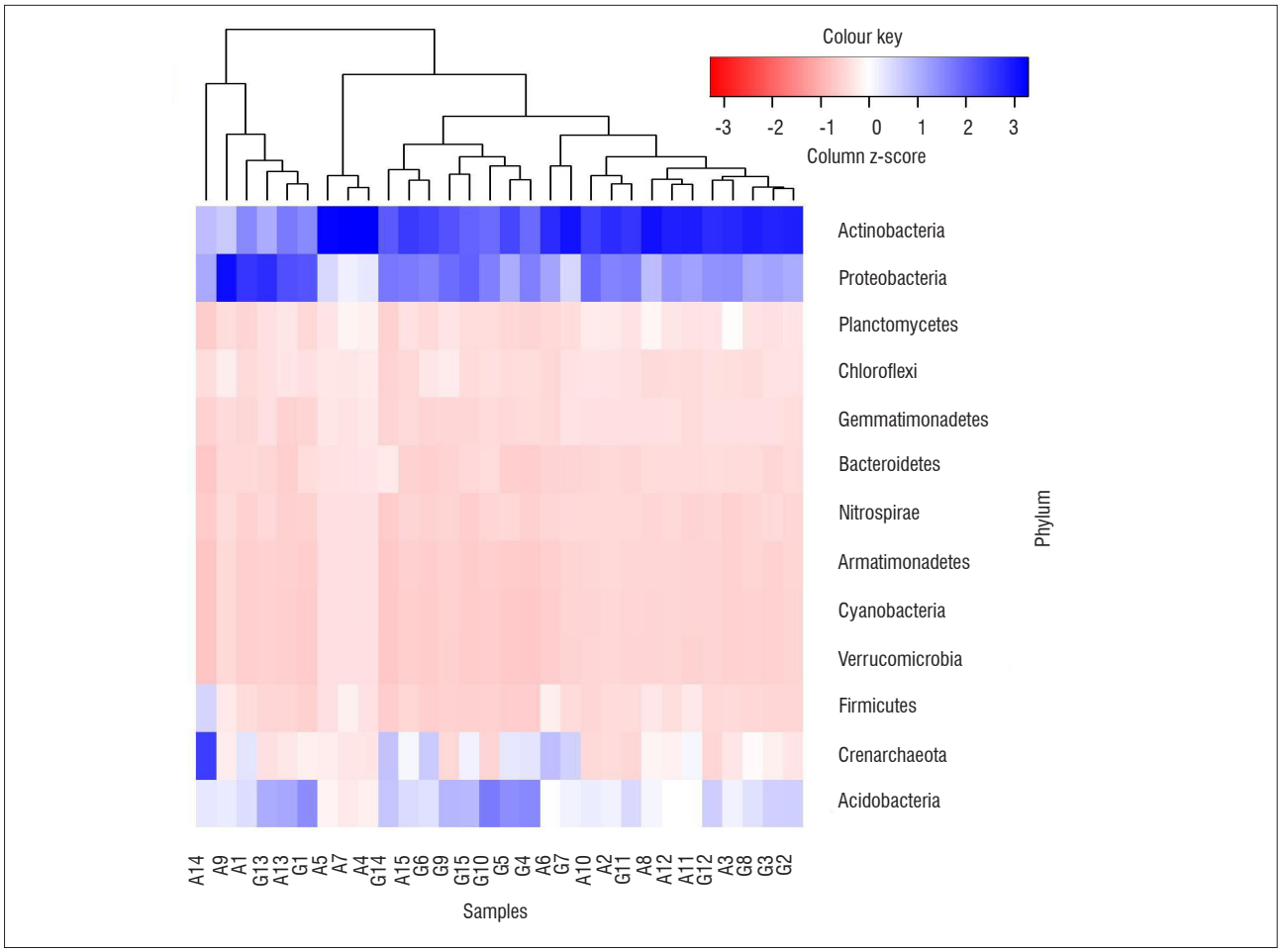


Figure 2: Heatmap displaying the bacteria and archaea phyla detected. Paired-site samples are clustered based on the per cent relative abundance. Each row was scaled so that the mean of each taxonomic group across samples was calculated and coloured by the corresponding z-score of each cell. Sample names starting with A were obtained from agricultural soils while those starting with G were from grassland soils.

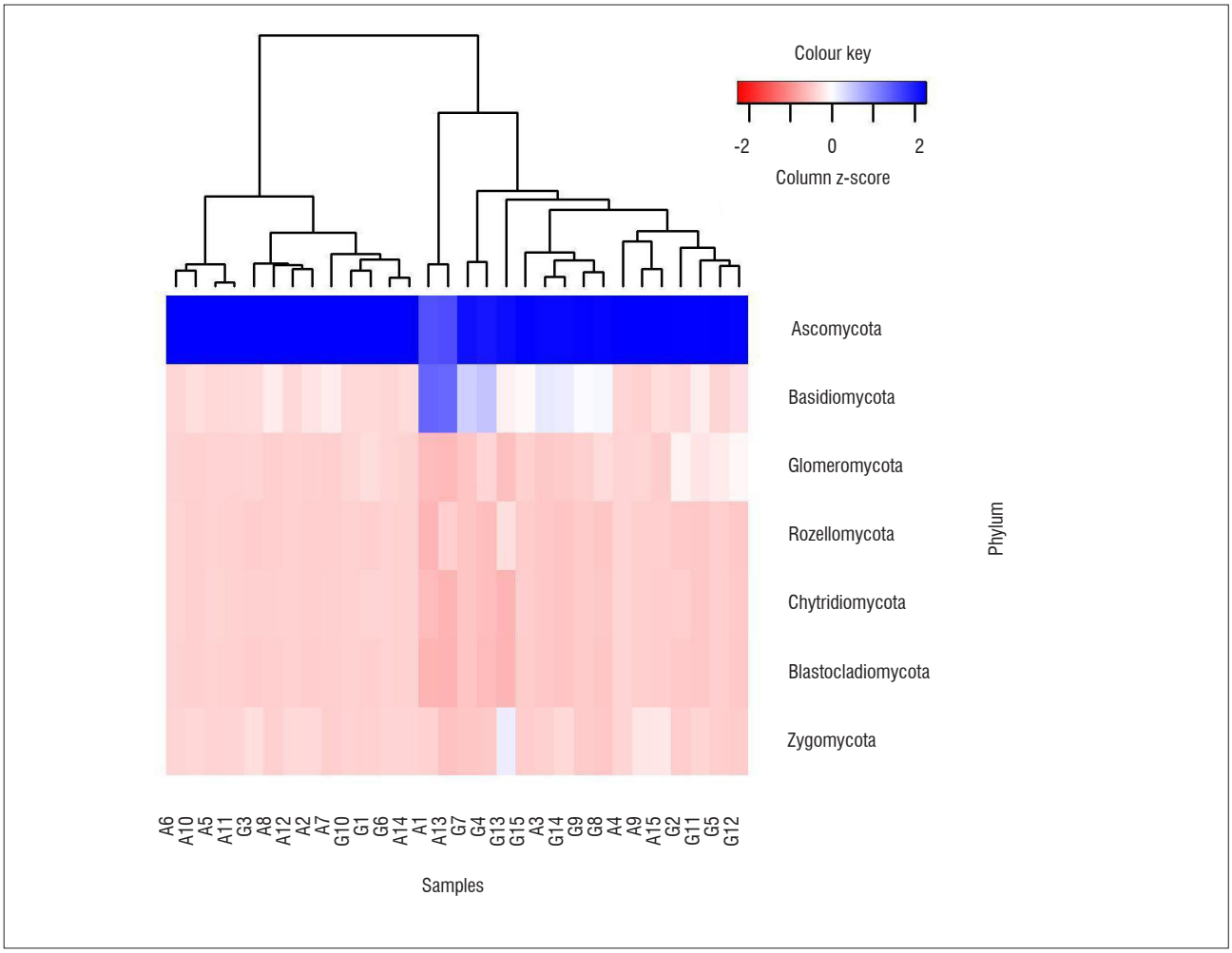


Figure 3: Heatmap displaying the fungal phyla detected. Paired-site samples are clustered based on the per cent relative abundance. Each row was scaled so that the mean of each taxonomic group across samples was calculated and coloured by the corresponding z-score of each cell. Sample names starting with A were obtained from agricultural soils while those starting with G were from grassland soils.

Non-metric multidimensional scaling analysis showed that bacterial and fungal communities in agricultural soils were distinct from those of the natural grassland soils (Supplementary figures 3 and 4). These results were confirmed by an analysis of similarity (ANOSIM: $R_{bacteria}=0.46$, $p=0.001$; $R_{fungi}=0.16$, $p=0.001$). A principal component analysis showed a clear discrimination of soil chemistry between agricultural soils and natural grassland soils. Soil pH and Ca concentrations tended to be higher in natural grassland soils, while NH_4, NO_3, K, C, P and Na concentrations were higher in agricultural soils (Supplementary figure 5). However, only three of the chemical parameters evaluated (Ca, P, pH) were significantly different (paired *t*-test, $p<0.05$) between the two habitats (Supplementary table 5). The high concentration of phosphorus observed in agricultural soils may be explained by mineral fertilisation commonly used in agriculture practices to improve plant performance.[55,56] Members of the soil microbial community, particularly bacteria, are capable of solubilising soil phosphate minerals into a usable form for plant uptake, which may influence the mineral concentration in the soil.[57] The differences observed in the composition and pattern of the bacterial and fungal communities (according to ANOSIM analyses) in agricultural soils compared with those in natural grassland soils, can potentially be explained by the differences in the soil chemistry of the two habitats. In previous studies on grassland soils, soil fertilisation and land-use intensification have both been linked to shifts in richness and diversity of soil microbial communities.[55,56]

Using distance-based redundancy analysis, we found that the composition and pattern of the bacterial communities in agricultural soils compared to natural grassland soils were influenced significantly by both soil pH ($R=0.008$, $p=0.002$) and phosphorus concentrations ($R=0.013$, $p=0.002$) (Supplementary figure 6). Our results are in agreement with previous studies which have shown that soil pH and nutrient availability shaped soil bacterial diversity.[2,9,13-15] In contrast, the differences found in the composition and pattern of the fungal communities between the two habitats were better explained by calcium ($R=0.019$, $p=0.01$) and phosphorus ($R=0.01$, $p=0.004$) concentrations (Supplementary figure 7), and corroborated previous studies.[9,58]

In summary, our results have shown that bacterial and fungal communities were significantly influenced by land-use changes, with agricultural soils containing distinct bacterial and fungal communities compared with natural grassland soils. Natural grassland soils consistently exhibited higher numbers of OTUs than did agricultural soils. The differences in microbial communities between the two habitat types were influenced by differences in soil chemical properties, supporting the concept that both soil chemical properties and microbial community compositions and diversities are altered after a conversion from a natural to an agricultural ecosystem.

This study was focused on a single economically important biome and represents the first demonstration of the effect of land-use changes on the diversity and structure of the soil microbial communities in African grassland soils. The study therefore serves as the benchmark for future studies on South African soil microbial diversity, and for monitoring future changes in soil microbial communities resulting from changing land use and climate.

Acknowledgements

We thank Dr Surendra Vikram and Mr Riegardt Johnson for providing bioinformatics support.

Authors' contributions

All authors conceived and designed the study and approved the final version; G.K.N. and A.V. collected the samples; G.K.N. performed research and analysed the data; G.K.N. wrote the manuscript; and D.A.C. and A.V. critically revised the manuscript.

References

1. Miransari M. Soil microbes and the availability of soil nutrients. Acta Physiol Plant. 2013;35:3075–3084. https://doi.org/10.1007/s11738-013-1338-2
2. Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci USA. 2006;103:626–631. https://doi.org/10.1073/pnas.0507535103
3. Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, et al. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. Ecol Lett. 2015;18:85–95. https://doi.org/10.1111/ele.12381
4. Singh BK, Bardgett RD, Smith P, Reay DS. Microorganisms and climate change: Terrestrial feedbacks and mitigation options. Nat Rev Microbiol. 2010;8:779–790. https://doi.org/10.1038/nrmicro2439
5. Schenck zu Schweinsberg-Mickan M, Jörgensen RG, Müller T. Rhizodeposition: Its contribution to microbial growth and carbon and nitrogen turnover within the rhizosphere. J Plant Nutr Soil Sci. 2012;175:750–760. https://doi.org/10.1002/jpln.201100300
6. Singh K, Trivedi P, Singh G, Singh B, Patra DD. Effect of different leaf litters on carbon, nitrogen and microbial activities of sodic soils. Land Degrad Dev. 2016;27:1215–1226. https://doi.org/10.1002/ldr.2313
7. Lehman RM, Cambardella CA, Stott DE, Acosta-Martinez V, Manter DK, Buyer JS, et al. Understanding and enhancing soil biological health: The solution for reversing soil degradation. Sustainability. 2015;7:988–1027. https://doi.org/10.3390/su7010988
8. Johnson MJ, Lee KY, Scow KM. DNA fingerprinting reveals links among agricultural crops, soil properties, and the composition of soil microbial communities. Geoderma. 2003;114:279–303. https://doi.org/10.1016/S0016-7061(03)00045-4
9. Lauber CL, Strickland MS, Bradford MA, Fierer N. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biol Biochem. 2008;40:2407–2415. https://doi.org/10.1016/j.soilbio.2008.05.021
10. García-Orenes F, Morugán-Coronado A, Zornoza R, Scow K. Changes in soil microbial community structure influenced by agricultural management practices in a Mediterranean Agro-Ecosystem. PLoS ONE. 2013;8, e80522, 9 pages. https://doi.org/10.1371/journal.pone.0080522
11. Lienhard P, Terrat S, Prévost-Bouré NC, Nowak V, Régnier T, Sayphoummie S, et al. Pyrosequencing evidences the impact of cropping on soil bacterial and fungal diversity in Laos tropical grassland. Agron Sustain Dev. 2014;34:525–533. https://doi.org/10.1007/s13593-013-0162-9
12. Acosta-Martínez V, Dowd S, Sun Y, Allen V. Tag-pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. Soil Biol Biochem. 2008;40:2762–2770. https://doi.org/10.1016/j.soilbio.2008.07.022
13. Lauber CL, Hamady M, Knight R, Fierer N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol. 2009;75:5111–5120. https://doi.org/10.1128/AEM.00335-09
14. Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, et al. Pyrosequencing-based assessment of bacterial community structure along different management types in German Forest and grassland soils. PLoS ONE. 2011;6, e17000, 12 pages. https://doi.org/10.1371/journal.pone.0017000
15. Kaiser K, Wemheuer B, Korolkow V, Wemheuer F, Nacke H, Schöning I, et al. Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. Sci Rep. 2016;6:33696. https://doi.org/10.1038/srep33696
16. Geremia RA, Puşcaş M, Zinger L, Bonneville JM, Choler P. Contrasting microbial biogeographical patterns between anthropogenic subalpine grasslands and natural alpine grasslands. New Phytol. 2016;209:1196–1207. https://doi.org/10.1111/nph.13690
17. Rutherford MC, Mucina L, Powrie LW. Biomes and bioregions of southern Africa. Strelitzia. 2006;19:32–50.
18. Bredenkamp G, Granger JE, Van Rooyen N. Moist Sandy Highveld Grassland. In: Low AB, Robelo AG, editors. Vegetation of South Africa, Lesotho and Swaziland. Pretoria: Department of Environmental Affairs and Tourism, 1996.
19. Abstract of Agricultural Statistic. Pretoria: Directorate: Statistics and Economic Analysis of the Department of Agriculture, Forestry and Fisheries; 2015.
20. Ferrar AA, Lötter MC. Mpumalanga biodiversity conservation plan handbook. Nelspruit: Mpumalanga Tourism and Parks Agency; 2007.
21. O'Connor TG, Bredenkamp GJ. Grassland. In: Cowling RM, Richardson DM, Pierce SM, editors. Vegetation of South Africa. Cambridge: Cambridge University Press; 1997.
22. Schofield RK, Taylor AW. The measurement of soil pH. Soil Sci Soc Am J. 1955;19:164–167. https://doi.org/10.2136/sssaj1955.03615995001900020013x
23. Bray RH, Kurtz LT. Determination of total, organic, and available forms of phosphorus in soils. Soil Sci. 1945;59:39–45. https://doi.org/10.1097/00010694-194501000-00006
24. Pansu M, Gautheyrou J. Handbook of soil analysis – Mineralogical, organic and inorganic methods. Berlin: Springer-Verlag; 2006. https://doi.org/10.1007/978-3-540-31211-6
25. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA. 2011;108:4516–4522. https://doi.org/10.1073/pnas.1000080107
26. White TJ, Bruns T, Lee S, Taylor J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninisky JJ, White TJ, editors. PCR protocols: A guide to methods and applications. New York: Academic Press; 1990. p. 315–322. https://doi.org/10.1016/B978-0-12-372180-8.50042-1
27. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–336. https://doi.org/10.1038/nmeth.f.303
28. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461
29. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72:5069–5072. https://doi.org/10.1128/AEM.03006-05
30. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013. Available from: http://www.R-project.org
31. Ramette A. Multivariate analyses in microbial ecology. FEMS Microbiol Ecol. 2007;62:142–160. https://doi.org/10.1111/j.1574-6941.2007.00375.x
32. Hassink J, Bouwman LA, Zwart KB, Brussaard L. Relationships between habitable pore space, soil biota and mineralization rates in grassland soils. Soil Biol Biochem. 1993;25:47–55. https://doi.org/10.1016/0038-0717(93)90240-C
33. Velvis H. Evaluation of the selective respiratory inhibition method for measuring the ratio of fungal: Bacterial activity in acid agricultural soils. Biol Fert Soils. 1997;25:354–360. https://doi.org/10.1007/s003740050325
34. Millard P, Singh BK. Does grassland vegetation drive soil microbial diversity? Nutr Cycl Agroecosys. 2010;88:147–158. https://doi.org/10.1007/s10705-009-9314-3
35. Mendes LW, Tsai SM, Navarrete AA, De Hollander M, Van Veen JA, Kuramae EE. Soil-borne microbiome: Linking diversity to function. Microbiol Ecol. 2015;70:255–265. https://doi.org/10.1007/s00248-014-0559-2

36. Frey SD, Elliott ET, Paustian K. Bacterial and fungal abundance and biomass in conventional and no-tillage agroecosystems along two climatic gradients. Soil Biol Biochem. 1999;31:573–585. https://doi.org/10.1016/S0038-0717(98)00161-8

37. De Vries FT, Hoffland E, Van Eekeren N, De Hollander M, Van Veen JA, Kuramae EE. Fungal/bacterial ratios in grasslands with contrasting nitrogen management. Soil Biol Biochem. 2006;38:2092–2103. https://doi.org/10.1016/j.soilbio.2006.01.008

38. Osanai Y, Bougoure DS, Hayden HL, Hovenden MJ. Co-occurring grass species differ in their associated microbial community composition in a temperate native grassland. Plant Soil. 2013;368:419–431. https://doi.org/10.1007/s11104-012-1529-4

39. Kowalchuk GA, Buma DS, De Boer W, Klinkhamer PGL, Van Veen JA. Effects of above-ground plant species composition and diversity on the diversity of soil-borne microorganisms. A Van Leeuw J Microb. 2002;81:509–520. https://doi.org/10.1023/A:1020565523615

40. Zak DR, Holmes WE, White DC, Peacock AD, Tilman D. Plant diversity, soil microbial communities, and ecosystem function: Are there any links? Ecology. 2003;84:2042–2050. https://doi.org/10.1890/02-0433

41. Berg G, Smalla K. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. FEMS Microbiol Ecol. 2009;68:1–13. https://doi.org/10.1111/j.1574-6941.2009.00654.x

42. Balesdent J, Chenu C, Balabane M. Relationship of soil organic matter dynamics to physical protection and tillage. Soil Till Res. 2000;53:215–230. https://doi.org/10.1016/S0167-1987(99)00107-5

43. Six J, Feller C, Denef K, Ogle SM, De Morae Sa JC, Albrecht A. Soil organic matter, biota and aggregation in temperate and tropical soils-effects of no-tillage. Agronomie. 2002;22:755–775. https://doi.org/10.1051/agro:2002043

44. Bloem J, Schouten T, Didden W, Jagers op Akkerhuis G, Keidel H, Rutgers M, et al. Measuring soil biodiversity: Experiences, impediments and research needs. In: Francaviglia R, editor. Proceedings of the OECD expert meeting on soil erosion and soil biodiversity indicators; 2004 March 25–28; Rome, Italy. Paris: OECD; 2004. p. 109–129.

45. Bardgett RD, Frankland JC, Whittaker JB. The effects of agricultural management on the soil biota of some upland grasslands. Agr Ecosyst Environ. 1993;45:25–45. https://doi.org/10.1016/0167-8809(93)90057-V

46. Bardgett RD, Hobbs PJ, Frostegard A. Changes in soil fungal: Bacterial biomass ratios following reductions in the intensity of management of an upland grassland. Biol Fert Soils. 1996;22:261–264. https://doi.org/10.1007/BF00382522

47. Janssen PH. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl Environ Microbiol. 2006;72:1719–1728. https://doi.org/10.1128/AEM.72.3.1719-1728.2006

48. Spain AM, Krumholz LR, Elshahed MS. Abundance, composition, diversity and novelty of soil Proteobacteria. ISME J. 2009;3:992–1000. https://doi.org/10.1038/ismej.2009.43

49. Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, et al. Global patterns in the biogeography of bacterial taxa. Environ Microbiol. 2011;13:135–144. https://doi.org/10.1111/j.1462-2920.2010.02315.x

50. Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, et al. Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. Appl Environ Microb. 2009;75:2046–2056. https://doi.org/10.1128/AEM.02294-08

51. Schlatter D, Fubuh A, Xiao K, Hernandez D, Hobbie S, Kinkel L. Resource amendments influence density and competitive phenotypes of *Streptomyces* in soil. Microbial Ecol. 2009;57:413–420. https://doi.org/10.1007/s00248-008-9433-4

52. Dharumadurai D, Jiang Y. Actinobacteria – Basics and biotechnological applications. Rijeka: InTech; 2016.

53. Marasas WFO, Nelson PE, Toussoun TA. Toxigenic *Fusarium* species: Identity and mycotoxicology. University Park, PA: Pennsylvania State University Press; 1984.

54. Rossman A, Seifert K. Phylogenetic revision of taxonomic concepts in the *Hypocreales* and other *Ascomycota*: A tribute to Gary J. Samuels. Stud Mycol. 2011;68:1–256.

55. Jangid K, Williams MA, Franzluebbers AJ, Sanderlin JS, Reeves JH, Jenkins MB, et al. Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. Soil Biol Biochem. 2008;40:2843–2853. https://doi.org/10.1016/j.soilbio.2008.07.030

56. Cassman NA, Leite MFA, Pan Y, De Hollander M, Van Veen JA, Kuramae EE. Plant and soil fungal but not soil bacterial communities are linked in long-term fertilized grassland. Sci Rep. 2016;6:23680. https://doi.org/10.1038/srep23680

57. Rodriguez H, Fraga R. Phosphate solubilizing bacteria and their role in plant growth promotion. Biotechnol Adv. 1999;17:319–339. https://doi.org/10.1016/S0734-9750(99)00014-2

58. Pec GJ, Karst J, Taylor DL, Cigan PW, Erbilgin N, Cooke JEK, et al. Change in soil fungal community structure driven by a decline in ectomycorrhizal fungi following a mountain pine beetle (*Dendroctonus ponderosae*) outbreak. New Phytol. 2017;213:864–873. https://doi.org/10.1111/nph.14195