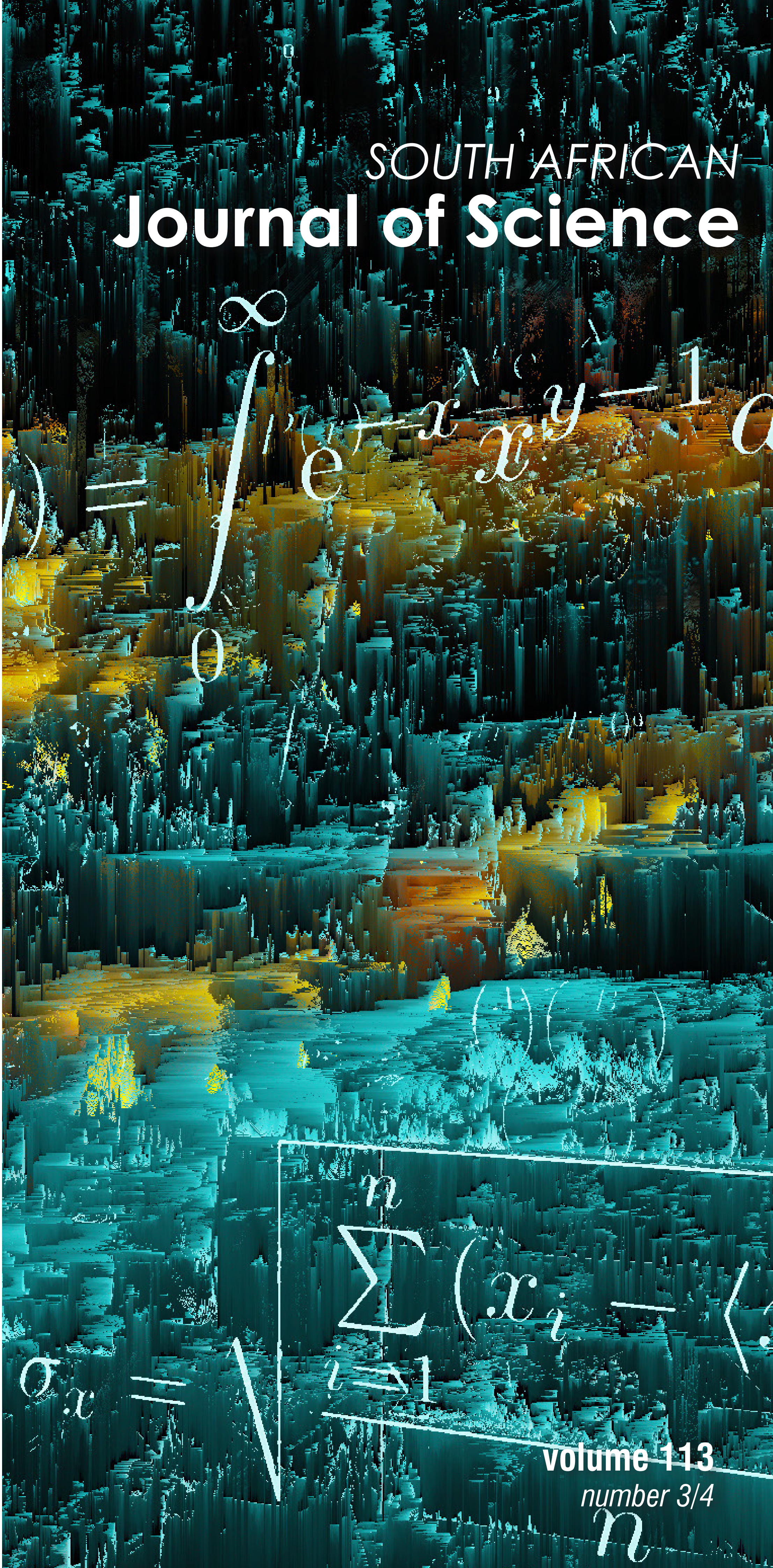
SOUTH AFRICAN
Journal of Science

Synergies in the mathematical sciences

Alternative age for Little Foot

Characterisation of traditional cosmetic clays

Determination of ancient arrow poisons



MARCH/APRIL 2017

eISSN: 1996-7489

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

volume 113
number 3/4

MARCH/APRIL

eISSN: 1996-7489

SOUTH AFRICAN Journal of Science

volume 113

number 3/4

Leader

Invited Commentary

Commentary

Book Review

Review Article

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

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

Correspondence and enquiries
sajs@assaf.org.za

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

Cover caption
The role of mathematics in different disciplines is explored in a series of Invited Commentaries on pages 2–16 (image CC: Torley, flickr).

Research Article

Mathematics and…

> *The vast majority of jobs of many flavours and incomes do not require the type of maths taught even in Grade 9. This is forgotten when mathematics is positioned as supremely important for the job market, or for students' personal development.*

This view, expressed in a recent article in *The Conversation* by Sara Muller, a researcher and doctoral candidate at the University of Cape Town, is not one generally shared.

Scores of articles in the media, written by other experts, have pointed to the dangers of lowering the pass mark for mathematics, and the importance of at least some level of literacy in mathematics. Clive Kronenberg of the Cape Peninsula University of Technology, for example, wrote:

> [Much research] *illustrates the debilitating burden that generations of South African children have had to endure, from the apartheid era until the present day: an education system that has failed them. It has not inducted pupils into the custom of thinking and reasoning on logical, rational and critical terms. Critical thinking is a vital skill.*
>
> *Research* has shown *that a well-cultivated critical thinker raises vital questions and problems, formulating them clearly and precisely. They are able to gather and assess relevant information, using abstract ideas to interpret it effectively. They can reach well-reasoned conclusions and solutions, testing them against relevant criteria and standards.*
>
> *The relationship between 'mathematics education' and 'more complex thinking' is typically symbiotic and mutually inclusive. Good, productive mathematics education can positively* raise pupils' skills *in diagnostic, methodical thinking.*

Ironically, it has also been pointed out recently that the levels of performance in the 'maths literacy' option are declining, just as is performance in the 'standard' mathematics courses. And, as is well known, South African schoolgoers perform close to the bottom of world measures of mathematics and language skills.

While it is probably true that there are many jobs that do not require reasonable levels of mathematical competence and the accompanying critical skills, many certainly do – and in some universities (the producers of high-level skilled employees), a surprising range of disciplines require some competence in mathematics or statistics. That students studying physics, chemistry, engineering, computer or information science, astronomy, oceanography and the animal and plant sciences need such competency is obvious. Increasingly, though, disciplines such as economics, palaeoanthropology, archaeology, psychology, education, geography and meteorology also require those skills.

Coming to terms with the value of mathematics in all varieties is not just a useful exercise then, but a critical part of education and of social and economic development. In view of the importance and significance of the mathematical sciences, the Academy of Science of South Africa (ASSAf) hosted a two-day workshop, in September 2016, entitled 'Finding Synergies in the Mathematical Sciences'. A primary aim of the workshop was to seek a discursive space to explore synergies in the mathematical sciences that could lead to interdisciplinary collaborations across various disciplines that include mathematics as a key pivot in their pursuits to understand, generate and use new knowledge. Explaining the importance of the workshop, ASSAf stated:

> [There] *is a global need for the mathematical sciences, which includes the discipline of mathematics, to work closer together to create discursive spaces to explore synergies and collaborations of mutual benefit. The mathematical sciences include all disciplines that use mathematics in their pursuits to understand, generate and use new knowledge. This includes disciplines from the natural sciences such as physics, biochemistry, and astronomy etc. to the social sciences such as economics, psychology, education, etc.*

The workshop included the presentation of five papers each on the role of mathematics in a different field, by leading scientists in those fields. Each of these speakers was invited to revise their presentation for publication in the Journal and we are very pleased to be able to publish them as a set of Invited Commentaries in this issue.

Professor Jill Adler has written on 'Mathematics in mathematics education'; Professor Jan-Hendrik Hofmeyr on 'Mathematics and biology'; Professor Loyiso Nongxa on 'Mathematical and statistical foundations and challenges of big data sciences'; Professor Daya Reddy on 'Finding synergies between the mathematical and physical sciences'; and Professor Martin Wittenberg on 'Mathematics and economics: We should expect better models'.

Readers of the Journal will find each of the papers insightful and, it is to be hoped, a means for encouraging others (especially young scientists) to extend their knowledge of and interests in the mathematical sciences. In his commentary, Professor Reddy reminds readers that

> [d]*epending on one's sources, the mathematical historical record dates back to around 1200 BCE. We are more familiar in the West with the seminal contributions of ancient Greece, although the mathematical heritage of the Islamic world from the 8th century onwards has become increasingly better known. Nevertheless one finds, in many different parts of the world, evidence of mathematical activity dating from antiquity. A further example relates to the Mayan civilization of central America: evidence of mathematical activity there appears to date back at least to 50 BCE.*

Mathematics is not just a discipline with a wide-ranging value, and a long history – but one that flourished, early on, in many parts of the world.

The proceedings report of the workshop is available here.

HOW TO CITE: Butler-Adam J. Mathematics and… S Afr J Sci. 2017;113(3/4), Art. #a0204, 1 page. http://dx.doi.org/10.17159/sajs.2017/a0204

Mathematics in mathematics education

AUTHOR:
Jill Adler[1]

AFFILIATION:
[1]Division of Mathematics Education, School of Education, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Jill Adler

EMAIL:
jill.adler@wits.ac.za

KEYWORDS:
mathematics teaching; teacher education; undergraduate curriculum; school mathematics curriculum

HOW TO CITE:
Adler J. Mathematics in mathematics education. S Afr J Sci. 2017;113(3/4), Art. #a0201, 3 pages. http://dx.doi.org/10.17159/sajs.2017/a0201

What role does mathematics play in the teaching and research of some or all of the ‘disciplines’ of mathematics education, science education and the social sciences? This question was that to which I was asked to respond at the Academy of Science of South Africa’s workshop on the mathematical sciences, held in September 2016. As suggested by the title, my focus here is on mathematics education, which is my field of expertise. An underlying interest or concern of participants in the workshop was what was offered in undergraduate mathematics at university, and how this did (or did not) support the ever-widening role mathematics is playing across disciplines. In addition, there were questions about the place and nature of undergraduate mathematics in the preparation for numerous diverse careers.

The workshop took place over 2 days, and mine was the last of five presentations, following that of mathematics in the earth and biological sciences, economical sciences, engineering and physical sciences, and the mathematical sciences itself. Each presentation was followed by discussion in smaller groups and then plenaries in which the different groups shared key points of discussion. Each of the first four presentations is also the focus of a separate Invited Commentary in this issue. Together they provide a wealth of insight into developments in research in engineering, biology, economics and mathematics, and the implications of these for an undergraduate curriculum in the mathematical sciences.

I was struck during the workshop discussions that the career of a mathematics teacher did not explicitly enter the landscape of possible careers emerging from studying mathematical sciences at university. Yet there was considerable attention to the limitations of the mathematical knowledge and mathematical ways of thinking of first-year students across these fields and, by implication, the quality of mathematics teaching in our secondary schools.

Teaching and research in mathematics teacher education was, appropriately, the planned focus of my talk. I opened my presentation with two critical questions that needed to be considered in the workshop:

1. Where in a discussion of the future of the mathematical sciences in a rapidly changing, challenging and exciting world do we locate the career of a future school mathematics teacher?
2. What does this location mean for mathematical sciences curricula or education at university?

The background document for the workshop was the 2025 review of the mathematical sciences in the USA.[1] I was curious whether and how preparation of teachers received attention in that review. Indeed there is a section, albeit small, on the importance of teachers for K–12 (kindergarten to Grade 12). It points in particular to countries that perform well in international mathematics assessments like TIMSS and PISA, through which top mathematics school-leavers enter, and might even have to compete for places in, degree programmes, ultimately leading to teaching careers. Perhaps implied in this is a criticism within the USA where, as in South Africa, teaching is not a high status profession. The point here, however, is that in the 2025 review there also is no mention on whether and how the mathematical preparation of teachers is or should be part of a consideration of the undergraduate curriculum in the mathematical sciences. As I think about this issue, I wonder about a similar review of the mathematical sciences in high-performing countries. Would there be a consideration of ‘mathematics for teaching’ alongside ‘mathematics for biology’, ‘mathematics for finance’, and ‘mathematics for economics’ as had been discussed in the earlier presentations in the workshop? What would be different? What is taken for granted?

As I worked my presentation into this Invited Commentary a few months after the workshop, I was reminded of the recent presentation by the outgoing President of the Mathematical Association of America, Francis Su, entitled ‘Mathematics helps people flourish’. It was a wonderful talk – passionate and compelling. His presentation spiralled around a quotation by the French philosopher Simone Weil: ‘Every being cries out silently to be read differently’[2]. Using stories from an inmate in a high-security prison in the USA, who enjoyed and was studying mathematics, and from Weil herself, Su asks: When you think about who is capable of and who wants to do mathematics, who do you think about? As I move into this commentary I wish to pose a similar question to you, the reader: Who do you think about when you consider who is capable of and who wants to be a mathematics teacher? Would you encourage talented undergraduate mathematics students into the profession of teaching?

These are not trivial questions – they are reflections of how knowledge and status works in society, and I thus do not suggest there are simple answers. But, as in the other fields, there is a great deal that we now know from research in mathematics education and from rigorous study of aspects of mathematics that have arisen and been driven by problems in mathematics education. How might this research inform answers to the questions I have posed above?

What do we know from research in mathematics education?

There is increasing agreement that there is a specificity to the mathematical knowledge that is required and used in the work of teaching, with the implication that this kind of knowledge should be included in pre-service and continuing teacher education. Substantial and influential research on this topic has been carried out at both primary[3] and secondary[4] levels, with both works building on Shulman’s[5] seminal work on the distinctive and significant nature of professional knowledge for teaching, particularly what he termed pedagogic content knowledge.

These results and other developments related to mathematics teachers’ professional knowledge are described in a recent review of relevant research.[6] For example, large-scale research studies have found a moderate association between teachers who have appropriate knowledge of mathematics and pedagogical training and improved teaching and learning. Also, if a teacher’s advanced mathematical ability exceeds a certain threshold, it produces negligible improvements in learner outputs. Furthermore, in terms of the usefulness of what is being taught, it has been found that calculus coursework enhances learner achievement in algebra but not necessarily geometry.

The authors also discuss weaknesses in the field. There is a lack of agreement about definitions and the language used to describe the specificity of mathematical knowledge for mathematics teaching and its basic concepts. There is also a lack of consensus on the boundaries between this knowledge, mathematics itself and mathematics pedagogy, and consequently on the best practices that can be adopted in the education of primary and secondary school maths teachers. The review, of course, encompasses far more than I have reflected here, but the points made are sufficient for the purposes of this commentary.

What do we learn from research in mathematics education?

The empirical evidence we have begs our attention. A frequently expressed view in South Africa, one that was stated in the discussion groups in the workshop, is that a prospective secondary mathematics teacher needs at least to have succeeded in second-year level pure mathematics at university. What then of the evidence that advanced courses in calculus do not support quality teaching of geometry? Is there support for school Geometry in the first two years of an undergraduate pure mathematics degree programme?

We need to also consider the empirical evidence of improved teaching and learning in mathematics being a function of appropriate levels of mathematics *and* pedagogic training. In any preparation for future teaching careers there thus needs to be appropriate attention to various areas of the school mathematics curriculum and whether and how topics across the undergraduate mathematical sciences courses span them all. In addition, degrees for mathematics teacher preparation also need to attend to pedagogical training.

Third is the overall result that teaching mathematics requires specialised ways of knowing mathematics. However, while this might mean different types and thresholds of mathematical knowledge for teachers at various levels from pre-primary to tertiary education, to date there are not clear descriptions of what these types and thresholds are and so what is ‘appropriate’ at the various levels.

What are current practices?

The present routes into secondary teaching mathematics are: a bachelor’s degree in (or at least with some) mathematics taught by mathematics departments at university, followed by a Postgraduate Certificate in Education (PGCE) taught by education departments; or a Bachelor of Education degree, with mathematics and education courses taught predominantly in education departments. Both models have their constraints. The former path, in which all specialised mathematical knowledge for teaching is condensed into 1 year in the PGCE, may provide insufficient knowledge of and for teaching geometry, statistics, probability and financial maths, which are part of the school curriculum, but not necessarily the undergraduate mathematics curriculum. As already noted, research has shown that the calculus taught at university does not support the teaching of geometry at school as it does algebra. It is likely the same holds for probability, statistics and financial mathematics, again posing questions for the undergraduate curriculum in the mathematical sciences if it is to provide the appropriate mathematical education of future teachers.

The mathematics taught as part of the BEd has different limitations. If we focus here only on the BEd degree for secondary mathematics teachers, we need to understand that its intake of students is also different. Many students enter the degree having passed mathematics in Grade 12 and with sufficient points for entry, but with mathematical knowledge that is often ritualised. This means they are able to execute procedures, and with some skill, but not able to grasp the mathematical principles underlying these processes. There is thus a need to revisit school mathematics, and do so from an advanced perspective, deepening what prospective teachers know and understand about school mathematics. And this needs to be offered across all the domains that comprise the school curriculum. For example, if you want teachers to be able to think probabilistically, then you need to include in the mathematics programme a course in which the key ideas, concepts and processes that they need to know is offered. And this needs to be done while introducing and developing higher-level mathematics, and courses in relevant pedagogy. This is a far broader curriculum, just mathematically speaking, and thus cannot also provide for depth and extension across domains as in a dedicated 3-year degree in pure mathematics as it currently stands. While the description I offer here is informed by my knowledge of the BEd degree for secondary mathematics teachers at the University of the Witwatersrand, I think the situation will be similar in other universities in South Africa.

This brief description of different forms of pre-service teacher preparation currently offered reveals that the strengths of one route are reflected in the limitations of the other, thus challenging views that promote one or other of these routes as ‘the best’ preparation for teachers. Here too we confront a problem that does not have clear or immediately visible solutions, particularly in the current South African context in which the performance curve in Grade 12 mathematics is skewed towards poor results, with relatively few learners obtaining 60%.

As a means of thinking further about mathematics in teaching and thus about the mathematical preparation of future mathematics teachers at all levels of schooling, I offer an example of an intervention into secondary mathematics teachers education in South Africa – focused at the transition from Grades 9 to 10 – that I suggest gives practical meaning to the notion of mathematics in mathematics teaching at this mid-secondary level.

A current project with specific focus on mathematics for teaching

The Wits Maths Connect Secondary (WMCS) project at the University of the Witwatersrand is a linked research and development project seeking to improve the teaching and learning of mathematics in some secondary schools in one province in South Africa, through the professional development of mathematics teachers. The goals are twofold: to improve teachers’ mathematics knowledge for teaching and their teaching practices and to study whether and how the intervention impacts learners’ learning.[7]

The WMCS professional development programme includes a 16-day mathematics for teaching course focused on mathematics relevant to teaching across the Grades 9 to 10 transition. Teachers attend eight 2-day sessions over the course of a year. Three quarters of each 2-day session focuses on mathematics with the remaining quarter on strategies for teaching. Participating teachers are required to complete independent assignments on mathematics and on teaching mathematics in between each of the 2-day sessions. The course focuses on algebra and functions, with some attention given to geometry and trigonometry. The topics were chosen according to their relative importance within the curriculum, as well as for their potential to leverage learning gains.

In the schools we were working with in the WMCS project, we found many teachers who were teaching Grades 8 and 9 did not and could not teach beyond this level. Many were ‘out-of-field’ teachers, meaning that they were either specialists in subjects other than mathematics, or had trained perhaps as primary mathematics teachers. Many expressed a lack of confidence in their mathematical knowledge and our research on their teaching showed that their explanations lacked clear focus and coherence.

We thus designed, developed and implemented a course that revisits mathematics that might be considered ‘known’ to the participants. The ways in which this is done is to deepen, strengthen and extend this existing knowledge. For example, special cases are explored and aspects of mathematics that have been assumed as ‘known’ are problematised. Connections are made across topics and concepts in the curriculum. New mathematics is also included, to extend teachers beyond the curriculum and the grade level which they are teaching, and so into Grades 10–12 mathematics. The teachers work on their knowledge of key concepts as well as their fluency with relevant mathematical procedures, and both in a context in which mathematical inquiry of these is encouraged and supported through carefully designed mathematical tasks.

Strategies for teaching in the mathematics teaching component of the course focuses on key elements of mathematics teaching practices. I do not provide detail here and refer readers to articles related to this work.[8,9]

The impact of the professional development course has been studied by assessing the learning gains among a cohort of 609 pupils in five schools over an academic year (2013). The major result of this study is that learners taught by the teachers who had taken the course significantly outperformed the learners in the same schools taught by teachers who had not taken the course.[10] This was a pilot study, and the results are only indicative. Nevertheless, the implications are that enhancing teachers' mathematical knowledge for teaching can lead to improvements in learning.

The point of the discussion of this intervention and its research is that the course is organised around a selection of mathematics for teachers that are considered 'appropriate' for teachers and teaching at those grade levels. The course is deliberate in its own pedagogic strategies and creates possibilities for teachers to move from their ritualised to more elaborated mathematical knowledge. While the mathematics content selection and how it is organised for teachers' learning extends beyond the levels they teach (Grades 8 and 9), these do not reach levels of undergraduate mathematics.

Undergraduate mathematics with mathematics teacher education

The WMCS study is not focused at the senior secondary level (Grades 10–12). However, there are implications for the shape of a well-rounded undergraduate curriculum in the mathematical sciences if it is to include mathematics teacher education in its landscape. A mathematics for teaching course for teachers teaching Grades 11 and 12 mathematics would of course have a different selection of mathematics. I suggest, nevertheless, that similar principles should hold. Revisiting key concepts in secondary school mathematics is important for prospective teachers. This kind of mathematics study needs to be added to courses in advanced mathematics, to offer prospective teachers opportunities to learn new mathematical knowledge (e.g. calculus) and appreciate advances in the discipline. What then about advanced courses in geometry, probability, statistics and financial mathematics? These are all topics in the secondary curriculum, and all are part of the mathematical sciences, but not necessarily part of a considered degree programme suited to the needs of prospective teachers.

Broadening curricula by offering a wider range of mathematics topics inevitably has consequences for working within and across the mathematical sciences. And we have not yet included the necessary focus needed on pedagogical training. This tension between a broader curriculum and depth in disciplinary specialisation is one that will not be easily resolved, but it is one that needs attention. And then there are implications for who teaches all these different elements of professional knowledge for mathematics teachers, for example courses in revisiting school mathematics. Mathematicians? Mathematics teacher educators? The notion of mathematics in mathematics teaching as including specialised knowledge also suggests that we would need to nurture the identities and specific expertise of 'educators' who may foster the next generation of maths teachers. There is, indeed, much to do, in what is a collaborative task for those in both the mathematical sciences and mathematics education.

Acknowledgements

This work is based on the research supported by the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation (NRF) of South Africa (grant number 71218). Any opinion, finding and conclusion or recommendation expressed in this material is that of the author and the NRF does not accept any liability in this regard.

References

1. National Research Council of the National Academies. The mathematical sciences in 2025. Washington DC: The National Academies Press; 2013. https://doi.org/10.17226/15269
2. Francis S. Mathematics for human flourishing [homepage on the Internet]. c2017 [cited 2017 Feb 24]. Available from: https://mathyawp.wordpress.com/2017/01/08/mathematics-for-human-flourishing/
3. Ball DL, Thames MH, Phelps G. Content knowledge for teaching: What makes it special? J Teach Educ. 2008;59(5):389–407. https://doi.org/10.1177/0022487108324554
4. Baumert J, Kunter M, Blum W, Brunner M, Voss T, Jordan A, et al. Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. Am Educ Res J. 2010;47(1):133–180. https://doi.org/10.3102/0002831209345157
5. Shulman L. Knowledge and teaching: Foundations of a new reform. Harvard Educ Rev. 1987;57(1):1–22. https://doi.org/10.17763/haer.57.1.j463w79r56455411
6. Hoover M, Mosvold R, Ball D, Lai Y. Making progress on mathematical knowledge for teaching. The Mathematics Enthusiast. 2016;13:12–14.
7. Adler J. The research and development learning curve 2010-2014: Report of the Wits FRF Mathematics Chair and the Wits Maths Connect Secondary (WMCS) project [document on the Internet]. c2014 [cited 2017 Feb 24]. Available from: http://www.nrf.ac.za/sites/default/files/documents/Professor%20Adler%20CoP%20presentation%202014.pdf
8. Adler J, Venkat H. Teachers' mathematical discourse in instruction: Focus on examples and explanations. In: Rollnick M, Venkat H, Loughran J, Askew M, editors. Exploring content knowledge for teaching science and mathematics. London: Routledge; 2014. p. 132–146.
9. Adler J, Ronda E. Mathematical discourse in instruction matters. In: Adler J, Sfard A, editors. Research for educational change: Transforming researchers' insights into improvement in mathematics teaching and learning. Abingdon: Routledge; 2017. p. 64–81.
10. Pournara C, Hodgen J, Adler J, Pillay V. Can improving teachers' knowledge of mathematics lead to gains in learners' attainment in mathematics? S Afr J Educ. 2015;35(3):10. https://doi.org/10.15700/saje.v35n3a1083

Finding synergies between the mathematical and physical sciences

AUTHOR:
Daya Reddy[1]

AFFILIATION:
[1]Department of Mathematics and Applied Mathematics, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Daya Reddy

EMAIL:
daya.reddy@uct.ac.za

KEYWORDS:
mathematics; astronomy; engineering; interdependence; curricula

HOW TO CITE:
Reddy D. Finding synergies between the mathematical and physical sciences. S Afr J Sci. 2017;113(3/4), Art. #a0202, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0202

Depending on one's sources, the mathematical historical record dates back to around 1200 BCE.[1,2] We are more familiar in the West with the seminal contributions of ancient Greece, although the mathematical heritage of the Islamic world from the 8th century onwards has become increasingly better known. Nevertheless, one finds, in many different parts of the world, engagement of a mathematical nature dating from antiquity. A further example relates to the Mayan civilization of central America: evidence of mathematical activity there appears to date back at least to 50 BCE. Similar remarks may be made in respect of mathematical developments in sub-Saharan Africa.[3]

While much is known, there is uncertainty about attributions of some mathematical concepts, and also of the extent to which communities in different parts of the world communicated with each other on topics of a mathematical nature. Certainly the picture that emerges is one of developments of particular concepts in multiple locations – sometimes contemporaneous, often at different times – with ebbs and flows in creativity, yet also with examples of cross-pollination. Thus the picture of a society or region operating in complete isolation, laying claim to specific inventions and developments, and with these ideas possibly flowing, if at all, in one direction, is at odds with the admittedly incomplete record. An example would be the complicated history of the concept of zero, its symbolic representation, and its possibly multiple origins in various parts of Asia, as well as in the Mayan civilization.[1]

A striking and enduring feature of the development of mathematics is the central role played by the physical sciences, most notably astronomy, in motivating mathematical developments. Engineering also has played a significant role, spurred for example by military considerations.

The late 17th century onwards witnessed a great flowering in mathematics, inspired by the quest for knowledge in the physical sciences, with central figures including Isaac Newton, Gottfried Wilhelm Leibniz and Leonhard Euler on topics in mechanics and astronomy, the studies of Joseph Fourier on heat conduction, and of James Clerk Maxwell on electromagnetism. These topics in turn paved the way in the 20th century for the paradigm-shifting theories of relativity and quantum mechanics.

The intense relationship between mathematics and the physical sciences has not abated at all with the passage of the 20th century. Examples of groundbreaking achievements in physics that are underpinned by innovative mathematics include the work of astrophysicist Subramanyan Chandrasekhar, who won recognition, including a Nobel Prize in Physics in 1983, for his formulation of theories for the stability and evolution of stars, including those that subsequently undergo collapse into compact brilliant stars known as white dwarfs.

The second half of the 20th century has seen the emergence and maturation, in parallel with the advent of computers of ever-increasing power, of what is referred to as computational science, or scientific computing: that is, investigations in which advanced computing capabilities are used to understand and solve complex problems in the physical and biological sciences, engineering, social sciences, and a myriad of other areas. The essence of computational science is the development of algorithms based on mathematical models, turning these into computer code and other forms of software, and their use in simulation. It is now commonly accepted as a peer methodology alongside the traditional forms of investigation, namely experiment and theory. It is appropriate in considering the beginnings of scientific computing to mention the remarkable work of Ada Lovelace (1815–1852), among whose most important contributions was the first published description of a stepwise sequence of operations for solving certain mathematical problems. Her contributions included the visionary idea of a machine that could manipulate symbols in accordance with a set of rules, rather than simply calculate.

Mathematics remains as central to the physical sciences and engineering as ever. It provides the language and the avenues through which to develop models of physical phenomena which provide insights into the phenomena concerned, as well as predictive capabilities that may, for example, suggest particular directions in experimental studies. In the world of engineering such models are central to design – that is, the development of devices or structures to perform a specific function. Mathematical modelling furthermore is used in the optimal design of processes and components through procedures that allow for sequential improvement of designs, each time tweaking the previous attempt in a systematic way by modifying parameters such as the geometry or material composition.

But the relationship between mathematics and the physical sciences and engineering is much more than that. It is a symbiotic one, in the sense that problems in the physical sciences provide fertile ground for the development of new mathematical theories and techniques. The late 20th century has been witness to particularly striking examples of areas of physics such as quantum field theory giving rise to significant new insights in mathematics, and to remarkable and deep mathematical theories.[4] An acknowledgement of the two-way nature of these relationships is crucial to guiding the way in which we shape curricula, and design and pursue research programmes.

The interwoven nature of mathematics with the physical sciences is illustrated here through a description of the 'life cycle' of a problem arising in mechanics and materials science, and concerned with the behaviour of metals under various loading conditions. A researcher seeking to explore this problem from a mathematical perspective would begin by developing a model that captures in a mathematical sense the properties of the material. By combining these properties with well-established laws of physics such as conservation of mass and momentum, one obtains what is known as a system of partial differential equations, that is, a set of equations that describes the variation of quantities in position and time. These equations constitute the mathematical model in this instance.

The next step in the process is naturally that of seeking a solution; with this at our disposal we would be able to use the model to explore its predictive capabilities, and if relevant, engage in the process of design to meet a specified need. Here we inevitably hit an obstacle, for it is seldom the case that realistic models are amenable to exact solution. What to do? As a first step, the mathematical process continues by finding out as much as possible about the solution, even without being able to construct it. This is known as qualitative analysis, and is an extremely important part of the investigation.

The qualitative process then sets the scene for the next best thing, of finding an approximate solution: one that is not exact, but which is sufficiently close to meet our needs. It is at this point that the computer comes into play: the investigator develops techniques and associated algorithms, that is, a logical sequence of steps and calculations, which are translated into code, and through which the approximate solutions are generated. An important part of this step is validation: is the model a sufficiently reliable representation of the phenomenon? Then, we must undertake the process of verification: has the problem led to an approximation of sufficient accuracy? And, are we able to quantify the error in the approximation? The two processes of validation and verification are sometimes likened to asking (1) whether we have solved the right problem and (2) whether we have solved the problem right! Once these questions have been answered satisfactorily, the process of exploration or design can then begin in earnest.

The above description addresses through a simple example one direction in the two-way relationship between mathematics and the physical sciences. What about the other?

Well, it may be that the process of exploration yields phenomena that had not been anticipated, but had perhaps been observed in simulations, and which were therefore not accounted for in the qualitative analysis. For example, under certain conditions the metal sample may undergo deformation in which parts of its bulk slide internally relative to each other. It is precisely these kinds of phenomena that led, more than three decades ago, to the development of a rich mathematical theory of function spaces which are able to capture such behaviour. This was new mathematics, whose genesis was the study of the physical phenomenon of plastic or irreversible deformation.

This example illustrates the highly interdependent nature of mathematics and the physical sciences. It also addresses an apparent dichotomy, between pure and applied mathematics, that has bedevilled mathematics particularly since the 20th century. Such a distinction has been largely absent in the development and practice of mathematics, from antiquity through to the early 19th century: typically mathematicians would move effortlessly back and forth between what we today refer to as pure and applied mathematics. While the dichotomy has been given credence, particularly in the mid- to later 20th century, there is now a growing acceptance that it is a counterproductive, if at all valid, dichotomy.

Rather, a more appropriate metaphor is the description by American mathematician Robert Zimmer of mathematics as a fabric – a woven artefact that derives its strength from its interconnectedness, but which may be weakened by a tear or gap.[4] Another view of the corpus of mathematics is that of a grand edifice, a cathedral, with its practitioners working alone or in small groups on creating beautiful and functional components, and who would, when stepping back, be able to admire the totality of the creation.

Likewise, the organisational distinction between statistics and mathematics is unproductive and can hinder interdisciplinary work. In developing mathematical models, the aim is always to develop a model that is sufficiently realistic, yet not so complex as to be intractable, even computationally. There is an element of uncertainty in most physical processes and phenomena, which is not captured by purely deterministic models. In this context, uncertainty quantification has become a major area of endeavour, and one in which there is a healthy interaction between mathematicians, statisticians and probabilists.

While the physical sciences provide the example par excellence of a set of disciplines that have a hugely fertile relationship with mathematics, they are not unique in this regard. Indeed, similar relationships exist between mathematics and areas such as chemistry, biochemistry, molecular biology, demography and other social sciences. Economics is another example: consider the mathematicians who have been awarded the Nobel Memorial Prize in Economic Sciences!

As with any endeavour that crosses disciplinary boundaries, the key challenge, over and above that of defining and formulating the problem, is one of navigating these boundaries in such a way as to develop an intimate understanding of other disciplinary 'cultures': ways of communicating, methodologies, and of understanding what is important. Such cross-pollination requires dedicated time and energy, and the obstacles can sometimes be frustrating, but with patience it is a way of working that brings rich rewards.

As has always been the case, new times bring new opportunities. In this regard, the dramatic technological advances of the 20th century have played a major role in opening up new areas of enquiry. Mathematics, like all other disciplines, must be agile enough to be able to respond energetically to these new opportunities.

An example is the emergence of data science as a product of the digital revolution – a reference to the unprecedented explosion in the capacity to acquire, store, manipulate and transmit huge volumes of data. The emergence of big data, as it is referred to, opens the way to new and exciting challenges across many disciplines. These scientific opportunities lie in identifying and characterising previously unseen patterns and unsuspected relationships, being able to simulate highly complex system dynamics, and mapping complex states. Whatever the opportunities – in studying environmental change, climate forecasting or migration patterns, for example – these will require multidisciplinary approaches, and mathematics will necessarily occupy a central place in these collaborations. It is vital that our students and researchers be ready to grasp these opportunities.

What are the implications of these developments for curricula in mathematics? Continuous reflection is required around the question as to the characterisation of a well-rounded mathematics graduate in the modern era. It is essential that students be exposed not only to the manner in which mathematics is applied to other areas, but also to the ways in which its own development is influenced by progress in areas outside mathematics. Thus, the interactions between mathematics and other areas in the natural sciences and beyond, and the two-way nature of these interactions, should inform curriculum development in a direct way.

These considerations may lead to new courses, majors and partnerships with other disciplines. Computation, which can aid discovery and provide new insights as well as serve as a functional tool, should be an integral part of mathematics curricula. Likewise, the place of topics such as uncertainty quantification and its foundational elements in randomness and probability should be carefully evaluated.

A certain degree of flexibility is required so that mathematicians are able to discern new directions and opportunities, and are willing to ensure that these developments influence the structure and content of curricula. The fundamental attributes of mathematics such as conceptual and abstract thinking and deductive reasoning will always be present and represent essential skills for any number of careers. Equally important is the need to convey, whether through teaching at all levels or in sharing research ideas, the inherent beauty of mathematics and its cultural value.

References

1. Needham J. Science and civilization in China. Volume 3: Mathematics and the sciences of the heavens and the earth. Cambridge: Cambridge University Press; 1959.

2. Bourbaki N. Elements of the history of mathematics. New York: Springer-Verlag; 1994. http://dx.doi.org/10.1007/978-3-642-61693-8

3. Gerdes P. On mathematics in the history of sub-Saharan Africa. Hist Math. 1994;21:345–376. http://dx.doi.org/10.1006/hmat.1994.1029

4. National Research Council of the National Academies. The mathematical sciences in 2025. Washington DC: The National Academies Press; 2013. https://doi.org/10.17226/15269

Mathematics and biology

AUTHOR:
Jan-Hendrik S. Hofmeyr[1]

AFFILIATION:
[1]Centre for Complex Systems in Transition, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Jan-Hendrik Hofmeyr

EMAIL:
jhsh@sun.ac.za

KEYWORDS:
modelling relation; simulation; metric biology; relational biology

HOW TO CITE:
Hofmeyr J-HS. Mathematics and biology. S Afr J Sci. 2017;113(3/4), Art. #a0203, 3 pages. http://dx.doi.org/10.17159/sajs.2017/a0203

I am a biochemist turned theoretical systems biologist, who, by my definition, is someone who attempts to explain the emergence of systemic functional properties of the living cell that result from the interactions of its components. Mathematical modelling and computer simulation is the bread and butter of my everyday work. However, my formal training in mathematics was 1 year of 'mathematics for biologists', which certainly did not equip me with the tools I needed when, early on in my research career, I embarked on a theoretical and computational analysis of metabolic networks. To study the time-dependent and steady-state network behaviour I had to master the theory of non-linear differential equations and their numerical analysis; the framework of metabolic control analysis required me to learn linear algebra and the theory of matrices, and I had to teach myself computer programming. Lately I have become fascinated with self-fabrication and the functional organisation of cell processes and found that I needed to know about abstract algebra, category theory and the theory of formal systems and formal languages. Although this journey through mathematics has been exciting and rewarding, I do wish that I had had the appropriate training as part of my undergraduate studies. Nevertheless, as a biologist who uses mathematics to explore biological phenomena (in contrast to a biomathematician, who mines biology for interesting mathematical problems), I have learnt a lot about the relationship between biology and mathematics – a relationship which through its entire history has always been an uneasy one. Why this is so and why this may be changing are the questions I wish to explore here.

Whereas physics and mathematics have cultivated a rich synergy from which they have both clearly benefitted, most biologists have regarded mathematics to be of marginal or no interest to their work or, at best, to be useful for everyday numerical and statistical treatments of experimental data but nothing more. How many of us who teach in some field of biology have not encountered students and colleagues that readily concede that they chose to study biology because they feared or disliked mathematics? Often these sentiments are expressed not as a resistance towards mathematics per se, but as against theory in general, as if it were possible to do theory-free experiments. This sad state of affairs is expressed in our tertiary-level biology curricula in which mathematics is usually relegated to a one-semester or two-semester first-year course, after which it is never to be heard of again. Even mathematics departments usually make short shrift of biomathematics; to the best of my knowledge Stellenbosch University is at present the only South African university to offer a biomathematics degree programme.

Biologists' disregard of mathematics is even more unfortunate in the light of the essential role that mathematical formalisms fulfil in all of science. Science rests on the bedrock of natural law – the twin beliefs that there are causal entailments in the natural phenomena that we study and that they can be imaged in inferential entailments in our formal systems. This act of imaging causality in terms of inference is what we call 'modelling'. When the theoretical biologist Robert Rosen wrote 'I have been, and remain, entirely committed to the idea that modelling is the essence of science and the habitat of all epistemology'[1], he emphasised his view that science is the art of establishing what he called 'modelling relations' between the natural world and the world of our formalisms. This view is one I fully share and one that has afforded me deep insight into what I do as a scientist.

What is a modelling relation? According to Rosen[2,3], a modelling relation aims to establish congruence between two systems – more specifically, between the elements of each system and between the entailment structures of each system. By accomplishing both these aspects, the orderly nature of one system is made to correspond to another system to the extent that the two systems have a degree of correspondence. Let me unpack this definition. A modelling relation establishes a relation between that part of the natural world we choose as our object of study (let us call it 'natural system N') and a 'formal system F' – the inferential entailment structure i of which mimics the causal entailment structure c in the natural system (Figure 1). This is done by choosing a set of observables that we believe characterises the natural system, and then constructing an *encoding* dictionary ε that maps observables in the natural system to input variables in the formal system. The inferential rules for manipulating the entities in F are then supposed to image the causal relations in N. What we hope to achieve is to bring the entailment structures in the two systems into alignment, so that, given an input set, the result of the inferential process in F can be *decoded* via mapping δ into the N to make a prediction about its behaviour. When our prediction matches the behaviour of the natural system, i.e. when the result of natural processes c matches the result of tracing the arrows through $\varepsilon \rightarrow i \rightarrow \delta$ we can claim that F is a model of a particular aspect of N and that that aspect of N is a realisation of F. If this is achieved, the modelling relationship is said to satisfy a commutativity condition. In the language of category theory, a modelling relation establishes a functorial correspondence between N and F, i.e. objects in N map to objects in F and mappings in N map to mappings in F (Figure 2a).

The modelling relation also helps to clarify the difference between a 'model' and a 'simulation'. The mapping g in Figure 2b is a simulation of f instead of a model of f; it differs from mapping $\varepsilon(f)$ in Figure 2a in that it is not an encoding of a mapping f in N into one in F. This implies that g does not model the causal entailment structure of f. More generally, i in Figure 1 is a simulation of c if no attempt has been made to bring inferential entailment in F into correspondence with causal entailment in N. A typical instance of a simulation is when a set of input-output data is fitted with a polynomial to generate the input-output curve. Such a simulation provides no insight into the causal structure of the system that generated the data.

With this understanding of what modelling entails, we can ask which formalisms mathematical biology has employed to model biological systems and phenomena. For biology the 20th century was very much the age of analysis in which the quantitative approach ruled. I call this the age of metric biology. In metric biology, all observables are quantifiable state variables and the concept of a succession of state transitions forms the basis of the description of the time-dependent behaviour of biological systems. Mathematical biology took its cue from mathematical physics; the vast majority of mathematical models of biological processes were and still are cast in the language of dynamical systems theory, using deterministic non-linear ordinary differential equations, partial differential equations and difference equations. The last few decades have also seen the increasing incorporation of stochasticity into biological models, with the use of stochastic differential equations, the master equation and Markov processes. Hybrid models mix continuous/discrete and deterministic/stochastic differential equations.

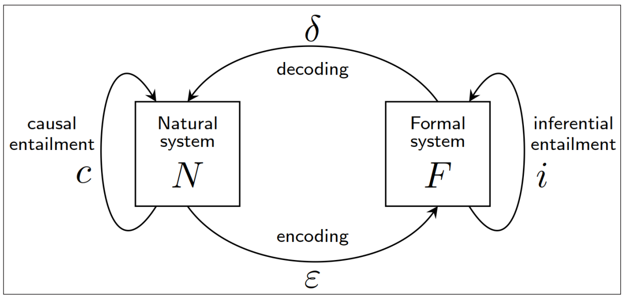


Figure 1: The modelling relation between a natural system *N* and a formal system *F*.

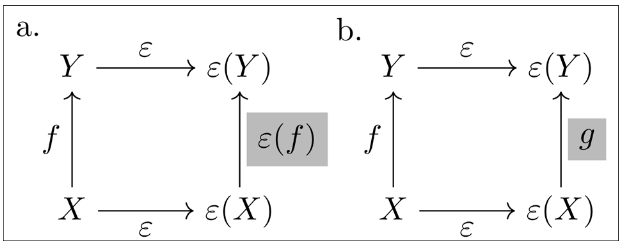


Figure 2: The difference between mappings between a natural system *N* and a formal system *F* in a (a) model and (b) simulation.

The major problem with the formal systems used to model metric biology is that they are impoverished in the type of causal entailment that they can image. Almost everything of interest in a system of ordinary differential equations is unentailed from within the system and must be posited from outside; all that remains internally is the transition from one state to the next.[3] If one asks the Aristotelean question 'why state x_{n+1}?' the only answer from within the system is 'because state x_n', which amounts to an answer in terms of material causation. Efficient causation (the mappings) and formal causation (initial conditions and parameters) are posited from outside and are immutable with regard to internal changes. Final causation (function) does not exist in this picture. But, in living systems, all these causes can be produced or altered from within the system. Consider, for example, an enzyme that catalyses a biochemical reaction and is therefore its efficient cause. In the cell, this enzyme is itself produced by ribosomal protein synthesis (its efficient cause) informed by the sequence of nucleotides in its messenger RNA (its formal cause); the production of the enzyme is the final cause (the biological function) of ribosomal protein synthesis. From a mathematical point of view the enzyme is simultaneously an operator and also an object in the range of another operator. This *functional entailment* is a hallmark of living systems.[3] In the sense discussed above, systems of differential equations simulate the behaviour of biological systems; they are not models because they do not capture the rich causal entailment structure of the biological system. Of course this does not imply that these simulations are not useful. If that were so I would have to discard three decades of my own systems biological research in which the behaviour of enzyme-catalysed networks are all described in terms of differential equations and through which I have learnt a lot about how these systems are controlled and regulated. Their main weakness is that they cannot capture the essence of living systems.

Nowhere is this realisation that we need new mathematics to model life more evident than in the research career of the father of mathematical biology. Nicholas Rashevsky originally trained as a theoretical physicist but turned to the study of biological phenomena after emigrating to the USA from the Ukraine. He started the school of mathematical biophysics at the University of Chicago in the 1930s, and in 1939 he founded the *Bulletin of Mathematical Biophysics*, the international journal of mathematical biology, re-titled in 1973 to the *Bulletin of Mathematical Biology.* Although there were of course earlier landmarks in the development of 20th century mathematical biology, such as D'Arcy Thompson's *On Growth and Form* and Alfred Lotka and Vito Volterra's predator–prey model, it is generally accepted that Rashevsky's work forms the bedrock of mathematical and theoretical biology. Taking his cue from mathematical physics, in his early work he used its methods to make fundamental contributions to mathematical biology: he studied bifurcation phenomena; a Boolean version of his 'two-factor theory' for excitable elements led in the hands of Walter Pitts and Warren McCullough to the development of neural networks, and was also a forerunner of Hodgkin and Huxley's model for the propagation of action potentials in neurons.

In the 1950s, Rashevsky began worrying that the mathematical tools that he had employed so successfully could not answer the most basic question of biology: What is life? In his own words:

> *All the theories mentioned above deal with separate biological phenomena. There is no record of a successful mathematical theory which would treat the integrated activities of the organism as a whole… And yet this integrated activity of the organism is probably the most essential manifestation of life… The fundamental manifestations of life mentioned above drop out from all our theories of mathematical biology… As we have seen, a direct application of the physical principles, used in the mathematical models of biological phenomena, for the purpose of building a theory of life as an aggregate of individual cells, is not likely to be fruitful. We must look for a principle which connects the different physical phenomena involved and expresses the biological unity of the organism and of the organic world as a whole.*[4]

It was this realisation that drove Rashevsky away from simulating state transitions to trying to model the functional organisation of processes in the organism. The emphasis shifted from the components of biological systems to the relations between them, and so was born what is now called 'relational biology'. He used topology, set theory ('organismic sets') and propositional logic as his new tools. Relational biology was more fully developed by his PhD student, Robert Rosen, who used category theory to formulate his so-called metabolism-repair systems, which captured the circular causal structure of cell processes that he termed closed to efficient causation. Rosen is now recognised as one of the most important theoretical biologists of the 20th century. His development of relational biology can be said to have heralded the advent of the age of synthesis, of an integrative view in which we try to glue Humpty Dumpty together again. If the age of analysis was characterised by the unspoken motto 'divide and conquer', then perhaps the age of synthesis aims to 'integrate and rule'. Today relational biology uses a rich array of mathematical tools: category theory, graph theory, network theory, automata theory, formal systems, to name the most important.

The perceptive reader may complain that I have not commented on the role of mathematics in fields such as evolutionary biology, population genetics, ecology and epidemiology. The reason is that, besides the crucial role of statistics in these disciplines, the mainstay has been population dynamics, which uses the mathematical tools of metric biology and therefore brings nothing new to the table.

As to the future of mathematical biology, Avner Friedman, who directed the Mathematical Biosciences Institute from 2002 to 2008, made this prediction:

> *Viewing the present trends in mathematical biology, I believe that the coming decade will demonstrate very clearly that mathematics is the future frontier of biology and biology is the future frontier of mathematics.*[5]

I agree with Friedman: my reading is that biology in general has finally embraced mathematics as an indispensable partner, and that biology enriches and will continue to enrich mathematics. It is now up to us to ensure that our biology and mathematics curricula are commensurate with this vision.

References

1. Rosen R. Essays on life itself. New York: Columbia University Press; 2000.
2. Rosen R. Anticipatory systems: Philosophical, mathematical and methodological foundations. 2nd ed. New York: Springer; 2012. https://doi.org/10.1007/978-1-4614-1269-4
3. Rosen R. Life Itself: A comprehensive inquiry into the nature, origin, and fabrication of life. New York: Columbia University Press; 1991.
4. Rashevsky N. Topology and life: In search of general mathematical principles in biology and sociology. Bull Math Biophys. 1954;6:317–348. https://doi.org/10.1007/BF02484495
5. Friedman A. What is mathematical biology and how useful is it? Notices of the AMS. 2010;August:851–857.

Mathematics and economics: We should expect better models

AUTHOR:
Martin Wittenberg[1]

AFFILIATION:
[1]School of Economics, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Martin Wittenberg

EMAIL:
martin.wittenberg@uct.ac.za

KEYWORDS:
microeconomics; macroeconomics; applied research; teaching; South Africa

HOW TO CITE:
Wittenberg M. Mathematics and economics: We should expect better models. S Afr J Sci. 2017;113(3/4), Art. #a0205, 3 pages. http://dx.doi.org/10.17159/sajs.2017/a0205

Mathematics is central to the modern discipline of economics, although there has always been some dispute about the value of this relationship. Leijonhufvud[1], in his satire of the economics profession, declared that the 'Math-Econ'(mathematical economists) appeared to be the high priests of this tribe. There was debate on whether their work was 'best to be regarded as religious, folklore and mythology, philosophical and 'scientific', or as sports and games'[1(p.334)].

In this commentary I want to consider whether the role of mathematics goes beyond 'sports and games' in economics and reflect on some of the ways in which mathematics is used in economics in general and in South African economics in particular.

Mathematics in microeconomic theory

Economic theory falls into two broad domains: microeconomics, concerned largely with the interactions of individuals and firms, and macroeconomics, concerned with the aggregate behaviour of economies. The dominant paradigm in microeconomics has been 'neoclassical' price theory culminating in the theory of general equilibrium. The approach is axiomatic: begin with assumptions about the key drivers of individual economic behaviour and then deduce the implications for aggregate outcomes. Two of the core assumptions underpinning this approach are:

1. Individuals have relatively fixed likes and dislikes ('preferences') which can be modelled by a 'utility function'.
2. Individuals are rational in the sense that they will act so as to realise their most preferred outcomes within the means available to them.

The mathematical machinery deployed is that of constrained optimisation with a dash of topology. In applying the framework to various specific markets, a number of innovations have been made. In many markets (e.g. agriculture and finance), the key issue is to model uncertainty. How do you decide what to plant or at what rate to sell if you need to foresee the state of the economy in the future? The approach is to marry the utility function to a probability distribution capturing the probability of different outcomes. The idea is that decisions will be governed by the expected value of the different outcomes. The framework for this approach was laid down by the mathematician John von Neumann and the economist Oskar Morgenstern and is hence referred to as 'von Neumann–Morgenstern utility theory'. It underpins the pricing models used in finance including the Black–Scholes equation. This is an intriguing case because here we see an economic theory feeding back into actual behaviour in markets that the theory is intended to describe.

The von Neumann–Morgenstern model of expected utility was first presented as an aside in their development of game theory.[2] The core purpose of that theory is to analyse strategic interactions, i.e. where the outcome achieved depends on the choices of all the agents involved. In these situations, agents need to foresee how their opponents are likely to react. One of the first applications of that theory outside economics was in modelling interactions between the USA and the Soviet Union in nuclear deterrence. Since then game theory has found applications in many different domains, and has even entered the general discourse. Concepts such as 'prisoner's dilemma' or a 'zero-sum game' have been applied to many negotiation contexts. Game theory has been particularly successful in elucidating the evolution of conflict and cooperation, even in the case of biology. Many fascinating interactions such as parent–child conflicts have been successfully modelled. From biology the concept of an 'evolutionarily stable strategy' has filtered back into economics.

One of the key areas in which this theory has been applied is in the study of 'institutions'.[3] The first economists were acutely aware that the economic well-being of countries depended on the right sort of laws and non-predatory behaviour by the state. During the elaboration of the neoclassical theory of general equilibrium, a concern with political institutions, legal frameworks and the norms underpinning market exchange were put to one side. More recently, these concerns have re-emerged as central to the performance of countries and regions.[4] Countries which have institutions that entrench predatory practices have much weaker growth performance than countries with more inclusive practices.

While modern microeconomics has a far richer understanding of the ways in which individuals interact there are still a number of continuities between the neoclassical price theory and some of the modern game theoretic models:

- The 'rational actor' model is common to both.
- Both tend to concentrate on predicting 'equilibrium' behaviour.
- The preference structures which underlie the behaviour are not themselves the subject of theoretical investigation – they are assumed to be pre-given and relatively unchanging.

All of these elements have come under scrutiny recently. The 'behavioural economics' research programme has investigated how individuals actually behave instead of starting from choice axioms describing how rational agents ought to choose.[5] This research has provided a number of robust insights into what motivates people. Firstly, many of these experiments show that human beings have a very poor intuitive grasp of probabilities, thus casting doubt on the ability of the 'expected utility' framework to adequately describe choice in uncertain conditions. Secondly, in laboratory experiments of various game theoretic interactions, such as the 'prisoner's dilemma' and the 'dictator' and 'ultimatum' games, it has been routinely found that individuals are more cooperative than the theory suggests they should be. Interestingly, students trained in economics tend to conform to the 'rational agent' model more closely than most people, suggesting once more that economic theory has the potential of influencing the actual behaviour of the agents it is supposed to describe.

A number of 'evolutionary' game theoretic models have explicitly considered less than perfectly rational agents learning from experience. These models also consider behaviour away from equilibrium. The resulting dynamics can be complicated and are typically simulated and not analytically solved out. These 'agent-based models' raise additional questions, because the space of all possible outcomes is obviously sampled rather than systematically described. Furthermore, the learning process is invariably local – it involves copying 'successful' behaviours from neighbours. It therefore differs from some of the theories to be discussed below which insist that agents should be able to learn some of the 'global' properties of the system in which they are embedded.

Finally a number of authors have begun to embed an analysis of human preferences and human behaviour in a broader account of the evolution of humanity.[6] This work draws attention to the culture in which agents operate and shows how this influences the types of behaviours that are possible.

Mathematics in macroeconomic theory

Macroeconomics has traditionally been the domain concerned with societal constraints. Its concerns have been with managing 'aggregate demand' and 'aggregate supply' so as to avoid the crippling depressions of the early 20th century. Post World War II, in the 'golden era' under the influence of Keynes, it appeared that this objective was being achieved.

The consensus was shattered by the sharp decline in growth of the 1970s. In this period the Keynesian models came under attack for not being based on proper 'microfoundations'. The mathematical theories which replaced them all tried to explicitly build up aggregate properties from microeconomic models. A key concept guiding this process was that of 'rational expectations'. The core idea behind this principle is that the market participants are forward looking and will be able to learn how the 'system' functions. So if, for example, the Reserve Bank continually injects money every time there is an economic downturn, people will eventually anticipate this behaviour and act in ways which then undercut the efficacy of that intervention. At a more conceptual level, the principle of 'rational expectations' can be thought of as an internal consistency check: if the model assumes that Y will always follow X, then the agents in the model should also be able to learn that Y will always follow X and act in accordance with that assumption.

One of the consequences of the 'rational expectations' revolution in macroeconomics was increasing scepticism about the effectiveness of policy more generally. If governments knew that doing X was going to cause Y, the rest of society would know this as well and would take steps to anticipate the outcome before X was even fully enacted. In market after market economists became convinced that government intervention was at best ineffective (if agents were free to act so as to undo what the government hoped to achieve) and at worst distortionary.

Some of the models built in adjustment costs which clawed back some role for policy, but on the whole the period since the 1980s saw a roll-back of the interventionism associated with the welfare state. The models that have recently been employed are called 'dynamic stochastic general equilibrium models'. They bolt together microeconomic general equilibrium (with a rational expectations twist) with growth processes (the 'dynamic' part) and random shocks.

The 2008 crash raised serious questions about the relevance of these models. A recent trenchant critique[7] suggested that there were several problems with the way in which these models were set up:

- The models all posit a 'representative consumer', i.e. all individuals are modelled as though they have identical preferences. The reality is that agents are heterogeneous and this heterogeneity matters for the way in which economic processes play out.
- The models assume that markets are perfect, that all agents have access to perfect information and that they can transact with no cost. However, real markets do not work like this. Indeed one of the big changes in microeconomics since the 1970s has been serious attention to the impact of these problems. For instance, in the absence of full information, not everyone who 'deserves' credit will be able to borrow. These problems will be exacerbated if contracts cannot be fully enforced, i.e. if it is not clear that individuals who could repay a loan will, in fact, do so. These constraints imply that many people will not get access to credit. They will thus save or hoard money for a bad day even if it would be theoretically more rational for them to spend it and, indeed, when the government is desperately trying to encourage spending to promote growth.
- The macroeconomic models also neglected the importance of the financial sector and in particular the specific institutional arrangements in different countries for dealing with debt.

Of course the reason why 'representative agent' models were adopted and transaction costs were ignored, is that these models are more tractable. The more heterogeneous the agents and the more imperfect the markets, the more difficult it is to model them – either analytically or even by simulation methods.

This short overview has highlighted several issues. Firstly, the set of mathematical tools available to economists has grown substantially over time. Nevertheless the success in properly describing or predicting behaviour is still far short of where it should be. Secondly, many economic theories have been developed with an eye on 'tractability', i.e. the ease with which they can be solved out, but in the process crucial features of the process to be modelled have been lost. Thirdly, economic theories have a way of 'looping back' into the material that they are supposed to be about. This can happen in two ways. Either agents absorb the theories and adapt their behaviour to align with the theory (e.g. when economic students use game theoretic concepts to work out how to behave in strategic interactions), or agents use the theory to anticipate the behaviour of other actors (e.g. the Reserve Bank) and potentially thereby undercut the outcome predicted by the theory. Modelling these self-referential loops in ways that go beyond the simple 'rational expectations' framework is one of the outstanding questions at the moment.

Finally, economic processes cannot be divorced from how 'institutions' in society operate, whether this be the way in which norms and laws either permit exploitation or foster cooperation, or the nature of the 'credit architecture' that channels borrowing or debts. A consideration of governance and politics and how this is incorporated into economic models is an important issue for theory.

Mathematics in applied research

It is safe to say that although there are lots of sophisticated economic models out there, their success in describing more than simple interactions (e.g. as captured by the 'prisoner's dilemma') is underwhelming. In taking models to 'the world', a number of additional tools derived from mathematics are used.

Econometrics is the branch of economics which marries mathematical models, statistics and economic theory. The workhorses of econometrics are similar to those of statistics – multiple regression analysis and related linear models. Nevertheless, there are some peculiarities to economic applications which have given econometrics a very particular flavour. One of the key issues is that the individuals from whom we collect the data to model behaviour (e.g. to estimate the determinants of wages) are not passive. Indeed, they may have an interest in the outcome of the research which can influence both whether they participate and what sort of information they are willing to volunteer. These 'sample selection' issues have been investigated for a while and econometricians have devised various techniques for dealing with them. Underlying all these are mathematical assumptions about how the response process works. Again these techniques have potential loops: it is difficult to understand who responds and how without reliable data, but the data cannot be 'corrected' to be made reliable without understanding the response process.

Another type of empirical model, used particularly in central banks and national treasuries, is the class of 'computable general equilibrium models'. These models attempt to analyse system-wide interactions. Typically they involve numerous 'sectors' (different industries) that buy from each other and sell output to 'households' and foreign countries, among others. These models are used to predict how particular 'shocks' will propagate through the economy. Unfortunately these models are quite opaque and the constituent equations are often not made available. Some of the equations in the model are estimated, but many of the key parameters are 'calibrated', i.e. they are selected by trial and error to get output that looks plausibly like that from the South African economy. This process is, of course, also not that transparent. The behaviour of the model depends on its structure, on the parameters that are chosen and how it is 'shocked'. The bigger the model the better it should be able to capture actual behaviour, but the more 'degrees of freedom' there are in tweaking parameters to get the desired output. It is not clear that the resulting predictions are worth the paper they are written on. The main function of such models seems to be to shut down debate about policy options rather than honestly explore the possibilities and the uncertainties associated with different choices.

Mathematics in teaching

Given the importance of models to much of economic theory and research, it stands to reason that trainee economists should be exposed to the mathematical techniques involved. Accordingly, economics departments often have a maths entrance requirement. In the case of the University of Cape Town, this requirement is actually quite high. Nevertheless, students struggle with the abstract nature of the models and as a result economics has in most standard undergraduate treatments replaced the mathematical relationships with graphical techniques. Consequently, first-year textbooks introduce many 'curves' (e.g. supply and demand or 'indifference curves') and then spend a lot of time discussing how these curves shift with changes in the underlying parameters of the model.

Students spend so much time trying to absorb the technical details that they often lose sight of the broader logical structure underpinning the model. Instead of understanding the model as a hypothetical construct (i.e. if the axioms of choice are true then we would expect to see the following behaviour), they view it as a set of statements of a priori truths, i.e. the models become quasi-theological statements which are impervious to disproof. This could be remedied to some extent if students saw alternative models earlier in their careers. There are now attempts under way to construct new syllabi that introduce transaction costs and non-clearing markets earlier in the sequence. Unfortunately some of these models make even higher mathematical demands.

And increasing the mathematical rigour of our teaching also imposes costs. For many students mathematics is seen as a straightforward barrier to entry. Economics and commerce courses are in big demand because they seem to offer a reliable route to higher paying jobs. Given the pressures to expand access, increasing the mathematical content may very well be seen as a retrograde step.

Mathematics in South African economics

Indeed at present it is not clear that for many South African economists the level of rigour is even strictly required. Very few South African economists work at the cutting edge of either economic theory or the application of new mathematical and statistical techniques to applied problems.

Nevertheless there are many opportunities for tackling some of the 'big' issues in South Africa. Modern economic theory has focused on the interplay between 'institutions' and economic processes. South Africa is arguably one of the better 'laboratories' in which to think about these connections. The history of apartheid and the process of transformation since 1994 provides a backdrop against which many of these issues can be researched, both empirically and theoretically. Of course a better modelling of these processes would undoubtedly loop back into the policy discussion itself. Getting the right concepts and tools for dealing with those questions could be enormously productive.

References

1. Leijonhufvud A. Life among the Econ. Econ Inq. 1973;11(3):327–337. https://doi.org/10.1111/j.1465-7295.1973.tb01065.x
2. Von Neumann J, Morgenstern O. Theory of games and economic behaviour. 3rd ed. Princeton: Princeton University Press; 1953.
3. Young HP. Individual strategy and social structure: An evolutionary theory of institutions. Princeton: Princeton University Press; 2001.
4. Acemoglu D, Robinson JA. Why nations fail: The origins of power, prosperity and poverty. London: Profile Books; 2012.
5. Kahneman D. Thinking fast and slow. New York: Farrar, Straus and Giroux; 2011.
6. Henrich J. The secret of our success. Princeton, NJ: Princeton University Press; 2016. https://doi.org/10.1515/9781400873296
7. Muellbauer J. Macroeconomics and consumption. Working Paper 811. Oxford: Economics Department, Oxford University; 2016. Available from: https://ideas.repec.org/p/oxf/wpaper/paper-811.html

Mathematical and statistical foundations and challenges of (big) data sciences

AUTHOR:
Loyiso G. Nongxa[1]

AFFILIATION:
[1]Centre for Mathematical and Computational Sciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Loyiso Nongxa

EMAIL:
Loyiso.nongxa@wits.ac.za

KEYWORDS:
high dimensionality; heterogeneity; data analytics; modelling; data sets

HOW TO CITE:
Nongxa LG. Mathematical and statistical foundations and challenges of (big) data sciences. S Afr J Sci. 2017;113(3/4), Art. #a0200, 4 pages. http://dx.doi.org/10.17159/sajs.2017/a0200

The hype around data sciences in general and big data in particular and the focus either on the potential commercial value of data analytics or on promoting its adoption as a new paradigm in conducting research, may crowd out important discussions that need to take place about the theoretical foundations of this 'emerging' discipline. In South Africa, discussions around (or the mere mention of) big data, especially within the National System of Innovation, often go hand in glove either with the Square Kilometre Array project and astrophysics, or eResearch or cyberinfrastructure. In his excellent essay '50 Years of data science', David Donoho of Stanford University remarks:

> *The now-contemplated field of data science amounts to a superset of the fields of statistics and machine learning which adds some technology for 'scaling' up to 'big data'. This chosen superset is motivated by commercial rather than intellectual developments. Choosing this way is likely to miss out on the really important intellectual development of the next fifty years.*[1]

It has now been recognised in some academic circles that important advances in the rapidly evolving 'discipline' of data sciences will depend significantly on contributions from communities in mathematics, statistics and computer science. This is reflected, amongst other things, by the number of conferences, workshops and thematic programmes that have been organised under themes like 'mathematical challenges of big data', or 'thematic programme on statistical inference, learning and models for big data'. Conversely, questions and open problems arising out of big data (will) spur new research activities and directions for the mathematical, statistical and computational sciences, and could lead to the opening of new frontiers in mathematics, statistics or theoretical computer science. The main objective in this commentary is to argue for the positioning of computational, mathematical and statistical sciences in South Africa at the centre of the heralded big data revolution. These disciplines are strategically important to provide a solid intellectual and academic foundation upon which to build a vibrant and successful project in big data analysis, especially in South Africa. This could lead to the renewal and rejuvenation of mathematical and statistical sciences in South Africa and facilitate collaboration and bringing closer researchers to work on problems at the boundaries or intersections of the respective disciplines. It also provides an opportunity to produce graduates with the breadth and depth of knowledge in all three major disciplines who are versatile enough to either be capable of pursuing fundamental research in these three broad disciplines or work in areas (public or private sectors) focusing on applications of (big) data sciences.

Background and terminology

The views expressed in this commentary are based on a desktop analysis of four different types of documents. The first are reports of workshops and thematic programmes funded and hosted by, for example, some National Science Foundation funded institutes like ICERM, SAMSI and IMA in the USA; the Fields and Banff Institutes in Canada; IMA in the United Kingdom and ACEMS in Australia. The second are abstracts of research papers presented at some of these workshops and survey articles in online research journals. The third type of documents considered are reports produced by Committees of the National Research Council of The National Academies – two worth highlighting are 'The Mathematical Sciences in 2025' and 'Frontiers in Massive Data Analysis'. And fourthly, syllabi and course outlines of mainly postgraduate courses that have been implemented at universities that have introduced new master's programmes in data sciences. In September 2016, the Academy of Science of South Africa convened a 2-day workshop under the theme 'Finding Synergies in the Mathematical Sciences'. This commentary is an overview of a background document prepared for a plenary presentation at the workshop.

There are terms that frequently crop up in discussions about (big) data analysis that often implicitly are used interchangeably as if they were synonymous. For the purposes of this commentary, we record the meaning that will be attached to these terms, although there is no consensus on the formal definition of these terms and how they are used. 'Data sciences' is an emerging 'science' concerned with extracting knowledge and information and gaining insights from data sets 'arising from experimental, observational, and/or simulated processes in the natural and social sciences and other areas'[2] which may be structured or unstructured and collected under diverse circumstances and environments. The 'theoretical foundations of data sciences' is a new area of academic interest at the intersection of statistical sciences and computer science, founded on a strong base of the mathematical sciences (pure and applied) and ranges from developing the theories, algorithms and methodologies, to paying attention to and understanding the applications of these within various domains. 'Data analytics' is the process of examining raw data with the purpose of knowledge discovery, gaining insights of information and hidden value from the data using software tools, techniques, processes and algorithms that have been specifically developed for such purposes. It is often argued that data scientists should have some specific domain knowledge, be it, for example in business management or health sciences. Data analytics in different domains gives rise to terms that signify the domain of the application: 'business analytics' is the application of data sciences in the business environment and has disproportionately received more attention, possibly more than applications in any other domain. One could (erroneously?) ascribe this phenomenon to the popularity of the book *Competing on Analytics: The New Science of Winning* by Jeanne G. Harris and Thomas H. Davenport and the highly cited and popular McKinsey Global Institute report, 'Big data: The next frontier for innovation, competition and productivity'. Other applications that have received significant attention would be health analytics, text analytics, fraud and risk analytics and financial analytics, to name just a few. On the other side of the spectrum, the influential book *The Fourth Paradigm: Data-intensive Scientific Discovery*, edited by Tony Hey, Stewart Tansley and Kirstin Tolle, makes the argument that scientific advances are becoming more and more data driven and researchers will more and more have to think of themselves as consumers of data. This has possibly influenced eResearch, defined by Whitmore[3] as:

> *the use of information technology to support existing and new forms of scholarly research in all academic disciplines…encompass*[ing] *computational and eScience, cyberinfrastructure and data curation…usually data-intensive but the concept also includes research performed digitally at any scale.*

Big data challenges:

There is an emerging consensus about the meaning of 'big data'[4-6], using the so-called four V's: volume, velocity, variety and veracity. Simply put, big data refers to massive data sets that are so voluminous and/or move so fast and/or vary in quality and structure, or may be 'questionable', which may make it difficult (to impossible) to store, manage and process using traditional methods. It is also generally accepted that one has to be mindful of the entire 'big data analysis pipeline' (sometimes called the big data analytics life cycle). This pipeline or life cycle involves two major phases: data management and data analytics. Each of these consists of clearly identifiable steps, each of which may present major challenges. Data management involves (1) acquisition and recording; (2) extraction, cleaning and annotation; and (3) integration, aggregation and representation. The analytics phase involves (1) data modelling and analysis; (2) interpretation; and (3) decision-making.

Some of the common challenges that underlie many, and sometimes all, of the different phases and processes above, require responses that depend significantly on ideas from the mathematical and statistical sciences.

Big data have unique features that are often not shared by or found in small or traditional data sets: (1) high dimensionality; (2) heterogeneity and incompleteness; (3) scale; (4) timeliness; and (5) security and privacy.

High dimensionality

A data point represents an object with, say p features, where p could be a very large positive integer. Geometrically the data point lives in a high-dimensional vector space. The geometry of high-dimensional vector spaces exhibits peculiarities which can be counter-intuitive when one attempts to extrapolate from lower-dimensional vector spaces and this is a major aspect of what is known as the 'curse of dimensionality'. For example, the volume of an n-ball tends to zero as n tends to infinity. If we have n data points, then in traditional statistical analysis, p may be fixed and much smaller than n (the data set can be represented by 'tall-and-skinny' matrices). But if p is much larger than n (data set represented by 'short and fat' matrices), or p increases as n increases, then the traditional methods of analysis could break down. It has been observed that high dimensionality could lead to spurious correlation, where variables that are (theoretically) independent may have high sample correlations. Spurious correlation may cause false scientific discoveries and wrong statistical inferences. The mathematical and statistical properties of high-dimensional data spaces are often poorly understood and inadequately considered.

Heterogeneity and incompleteness

One of the steps identified above in the big data analysis pipeline is the integration, aggregation and representation of data sets. These will be data sets corresponding to different populations that could have been collected from different sites and different environments, using different platforms, methodologies and technologies. Algorithms deployed on computational devices 'expect' homogeneous data and the systems have been designed to analyse data more efficiently if the data are structured and have identical format and size. A data set could have points with missing records and 'educated' estimates of these missing points could introduce biases and skew conclusions. Also omitting points with missing records could lead to different conclusions and recommendations.

Scale

The growth in computational and storage power has made it possible to work with massive data sets rather than small samples one normally deals with in traditional statistical analysis. It is common to discard or ignore outliers when analysing data sets of small size (which may be the case in traditional statistical analysis). However, these could in fact be representatives of what could turn out to be important subpopulations. Big data analysis allows for analyses that may reveal hidden structures of each subpopulation of the data. It may also reveal important common features across many subpopulations even when there are large individual variations. There are technical challenges that have been identified when it comes to working with data at peta-exascale. The 'fundamental shift underway now [is]: data volume is scaling faster than compute resources, and CPU speeds are static'[7].

Timeliness

The increase in the size of data sets to be analysed using the traditional methods and techniques means that it often takes longer to analyse. There could be situations in which there are time constraints to analysis, interpretation and decision-making and the result of the analysis is required immediately. This would be the case when a high-quality answer that is obtained slowly can be less useful (and more costly) than a medium-quality answer that is obtained quickly. An example is credit card fraud detection: it would bring down costs and prevent financial loss if this is flagged before a transaction is completed, potentially preventing the transaction from taking place. It is obvious that a full analysis of the background of the credit card owner's spending patterns may not be feasible in real-time and rather a decision would be based on a partial analysis.

Security and privacy

New capabilities to gather, analyse, disseminate, and preserve vast quantities of data raise new concerns about the nature of privacy and the means by which individual privacy might be compromised or protected.[8] There is justifiable public fear regarding the inappropriate uses of personal data. There is a conflict between the use of personal information for a 'greater' (?) public good on the one hand and disadvantaging individuals as a result of 'insights' gained from big data analysis on the other hand. This is not new and has occurred throughout our history and continues today. Today's concern about big data reflects both the substantial increase in the amount of data being collected and the fact that in many instances ordinary citizens are not even aware of the data that are being collected about them, nor aware of who has the data and how the data may be used against them. The sharing of data about individuals, without their consent, by those who hold such data and might be interested in aggregating the data to gain new insights is a big concern. The same data and analytics that provide benefits to individuals and society if used appropriately can also create potential harm. GPS tracking of individuals might lead to a better understanding of traffic problems which could lead to innovative solutions; but can also be used inappropriately in tracking the movement of individuals and their whereabouts. The recently published book by a former Wall Street quant, Cathy O'Neil, titled *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, raises some legitimate concerns about some of these issues.

Big data techniques

It was noted earlier that a data point can be viewed as an n-dimensional vector, where n represents the number of features that are being observed. If we consider a set of m such data points, then these can be represented as an m by n matrix. The data points can be thought of as being randomly selected from a probability distribution of this n-dimensional vector space. In certain instances, tensors, which are multidimensional generalisations of matrices, have been found to be useful in representing multidimensional data, for example in neurosciences. Data points can sometimes be viewed as 'nodes' of a graph, where the 'edges' represent, for example, 'relationships' between the nodes. The number of nodes and/or edges may increase or decrease with time and the edges may carry 'weights'. Topological methods can sometimes be deployed to determine the structural features of the data sets and geometry employed to determine the 'distance' between the data points. In some cases the data set can be too large to store or difficult to capture because it is arriving at high frequency with respect to the analysis resources available and sophisticated sampling and other statistical analysis techniques are required. The analysis of the data has to be as efficient and cost effective (in time, resources, etc.) as possible. This provides some indication as to why optimisation is ubiquitous in (big) data analytics.

Many of the big data techniques that often are deployed to efficiently process large volumes of data are founded on some of the mathematical and statistical concepts implicit in the previous paragraph and others. These include, but are not restricted to: (1) dimension reduction; (2) clustering; (3) data visualisation; (4) optimisation methods; (5) statistical methods (computing, inference, modelling and learning); (6) data mining; (7) machine learning; and (8) social network analysis.[9]

Mathematical and statistical challenges

The first issue to address is that of workforce development in an area that has been described as an emerging discipline and covers a broad range of learning areas that currently are taught in silos. The three communities – mathematics, statistics and computer science – in South Africa need to rise to the challenge of training the next generation of graduates who will have the necessary breadth and depth of knowledge in these three disciplines which are necessary to keep abreast of developments in the theoretical foundations of (big) data sciences and contribute to future developments of this new discipline. The graduates will also have to be comfortable working in multidisciplinary teams both in the public and private sectors. The core and indispensable knowledge areas include linear algebra, multivariable calculus, elementary probability and statistics, as well as experience in writing code in one of the main programming languages. These would normally be taught in the first 2 years at the undergraduate level. At the senior undergraduate level up to honours, matrix analysis, optimisation, graph theory and algebraic topology (emphasis on homology) would be foundational knowledge areas and other important topics would include harmonic analysis, time series analysis, approximation theory and a firm grasp of functional analysis. Stand-alone modules for each of these would lead to a 'bloated' curriculum and the challenge is for the experts to rethink the core knowledge and reorganise it into new packages that aggregate and integrate topics that are usually taught in silos. The more advanced topics are either not prominent or absent on the national research landscape in South Africa. In the mathematical sciences, these topics would include: (1) randomised numerical linear algebra; (2) topological data analysis; (3) matrix and tensor decompositions; (4) random graphs; (5) random matrices; and (6) complex networks.

Over the last 4 years, there has been a big focus on the foundations of (big) data sciences both in North America and the United Kingdom. There have been extended long-term thematic programmes hosted mainly (but not exclusively) by institutes in the USA funded by the National Science Foundation and privately, and government- and private-funded institutes in Canada. In the United Kingdom, the Engineering and Physical Sciences Research Council (EPSRC) has established (and is funding) Centres for Doctoral Training (CDTs) with a focus on various aspects (mainly statistical and computational) of data sciences. These offer a 4-year PhD with a compulsory large volume of coursework in the first year. Examples of thematic programmes are 'Theoretical Foundations of Big Data Analysis' hosted by the Simons Institute for the Theory of Computing (22/08/2013 – 20/12/2013); 'Statistical and Computational Methodology for Massive Data Sets', hosted by the Statistical and Mathematical Sciences Institute (2012/2013) and the 'Thematic Programme on Statistical Inference, Learning, and Models for Big Data' hosted by the Fields Institute for Research in Mathematical Sciences (06/2015 – 12/2015). The EPSRC funds at least three CDTs and their themes are 'Data Sciences' hosted by the University of Edinburgh, 'Cloud Computing for Big Data' hosted by the University of Newcastle and 'Next-generational Statistical Science' co-hosted by Oxford University and Warwick University. The reports of the thematic programmes and the coursework component of the CDT PhDs are a rich source of information both for topics for coursework in statistical sciences as well as open problems and research directions. The broad and common topics for coursework include (1) statistical modelling; (2) statistical and machine learning; (3) statistical inference; (4) computational statistics; (5) probability and approximation; (6) Bayesian methods for big data analysis; and (7) probabilistic methods for big data.

Future directions

As previously remarked, the reports of the thematic programmes hosted by national institutes provide a reservoir of information about, amongst other things, current research as well as research questions that would make a significant impact on and contribute to the theoretical foundations of data sciences. To provide a taste of some of these, below is an extract from a Call for Proposals by the US National Science Foundation under its 'Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Sciences and Engineering' programme:

- Sophisticated computational/statistical modelling for simulation, prediction, and assessment in computation-intensive and data-intensive scientific problems.
- State-of-the-art tools and theory in statistical inference and statistical learning for knowledge discovery from massive, complex, and dynamic data sets.
- General theory and algorithms for advancing large-scale modelling of problems that present particular computational difficulties, such as strong heterogeneities and anisotropies, multiphysics coupling, multiscale behaviour, stochastic forcing, uncertain parameters or dynamic data, and long-time behaviour.
- Study of mathematical, statistical, and stochastic properties of networks.
- Mathematical and statistical challenges of uncertainty quantification.
- Development of numerical, symbolic and statistical theory and tools to uncover and study analytical, topological, algebraic, geometric, and number-theoretic structures relevant for large-scale data acquisition, data security and cybersecurity.

Conclusion

The main objective of this short commentary was to propose to the mathematics, statistics and computer science communities at South African universities as well as private and public sectors to take an interest in the theoretical foundations of data sciences. This interest has potential to foster dialogue and collaboration amongst members of these communities. Such collaboration could spur rejuvenation and renewal in these three disciplines through incubating new areas of study on the South African higher education landscape and graduating the next generation of scholars with breadth and depth of knowledge in mathematics, statistics and computer science. Real and meaningful progress in big data does not only require 'whiz kids' in Hadoop, MapReduce, Spark, Python and R but also graduates who understand the algorithmic, mathematical and statistical underpinnings of these programs. The call we are making has potential for new collaborations between the university sector and users of mathematics, statistics and computer science in the private and public sectors and non-governmental organisations.

Acknowledgements

This commentary is based on a presentation I gave at a 2-day workshop organised by the Mathematical Sciences Committee of the Academy of Science of South Africa. I would like to thank Professor Fritz Hahne for bringing to my attention the report of the National Academy of Sciences titled 'The Mathematical Sciences in 2025' when I was on a fellowship at the Stellenbosch Institute for Advanced Studies (STIAS). I would also like to express my gratitude to STIAS for their generous financial support.

References

1. Donoho D. 50 Years of data science [document on the Internet]. c2015 [cited 2016 Jun 07]. Available from: http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf
2. Workshop on "Theoretical Foundations of Data Science (TFoDS)" [document on the Internet]. c2016 [cited 2017 Mar 09]. Available from: http://www.cs.rpi.edu/TFoDS/TFoDS_v5.pdf
3. Whitmore AL. Thoughts on eResearch: A scientist's perspective. J eScience Librariansh. 2013;2(2), e1045, 5 pages. http://dx.doi.org/10.7191/jeslib.2013.1045
4. Fan J, Han F, Liu H. Challenges of big data analysis. Natl Sci Rev. 2014;1(2):293–314. https://doi.org/10.1093/nsr/nwt032
5. Jagadish HV, Gehrke J, Labrinidis A, Papakonstantinou Y, Shahabi C. Big data and its technical challenges. Commun ACM. 2014;57(7):86–94. https://doi.org/10.1145/2611567
6. Agrawal D, Bernstein P, Bertino E, Davidson S, Dayal U, Franklin M, et al. Challenges and opportunities with big data: A white paper prepared for the Computing Community Consortium Committee of the Computing Research Association [document on the Internet]. c2012 [cited 2017 Mar 09]. Available from: http://cra.org/ccc/resources/ccc-led-whitepapers/
7. Lodha R, Jan H, Kurup L. Big data challenges: Data analysis perspective. Int J Current Eng Technol. 2014;4(5):3286–3289.
8. Moura J, Serrao C. Security and privacy issues of big data [article on the Internet]. c2016 [cited 2017 Jan 21]. Available from: https://arxiv.org/abs/1601.06206
9. Chen CLP, Zhang C-Y. Data-intensive applications, challenges, techniques and technologies: A survey on big data. Inform Sci. 2014;275:314–347. http://dx.doi.org/10.1016/j.ins.2014.01.015

Livestock predation in South Africa: The need for and value of a scientific assessment

AUTHORS:
Graham I.H. Kerley[1]
Kevin G. Behrens[2]
Jane Carruthers[3]
Marius Diemont[4]
Jurie du Plessis[5]
Liaan Minnie[1]
Philip R.K. Richardson[6]
Michael J. Somers[7]
Craig J. Tambling[8]
Jane Turpie[9]
Hermias N. van Niekerk[10]
Dave Balfour[11]

AFFILIATIONS:
[1]Centre for African Conservation Ecology, Nelson Mandela University, Port Elizabeth, South Africa
[2]Steve Biko Centre for Bioethics, University of the Witwatersrand, Johannesburg, South Africa
[3]University of South Africa, Pretoria, South Africa
[4]Webber Wentzel, Cape Town, South Africa
[5]Department of Mammalogy, National Museum, Bloemfontein, South Africa
[6]Human Wildlife Solutions, Cape Town, South Africa
[7]Centre for Wildlife Management, Centre for Invasion Biology, University of Pretoria, Pretoria, South Africa.
[8]Department of Zoology and Entomology, University of Fort Hare, Alice, South Africa
[9]Environmental Policy Research Unit, School of Economics, University of Cape Town, Cape Town, South Africa
[10]Unit for Livestock and Wildlife, Department of Agricultural Economics, University of the Free State, Bloemfontein, South Africa
[11]Independent Researcher, East London, South Africa

CORRESPONDENCE TO:
Graham Kerley

EMAIL:
graham.kerley@nmmu.ac.za

KEYWORDS:
evidence-based policy; legislation; transdisciplinary; wildlife

HOW TO CITE:
Kerley GIH, Behrens KG, Carruthers J, Diemont M, Du Plessis J, Minnie L, et al. Livestock predation in South Africa: The need for and value of a scientific assessment. S Afr J Sci. 2017;113(3/4), Art. #a0198, 3 pages. http://dx.doi.org/10.17159/sajs.2017/a0198

Predation of livestock in South Africa has been estimated to cost in excess of ZAR1 billion in losses per year[1] and has complex social, economic and ecological drivers and consequences. In this context, livestock can be broadly defined as domesticated animals and wildlife (the former excluding poultry and the latter including ostrich, *Struthio camelus*) managed for commercial purposes or human benefit in free-ranging (or semi-free ranging) circumstances that render them vulnerable to predation. This conflict between livestock producers and predators, and the attempts to manage it, has persisted for over 350 years, with the most notable outcome being the eradication of the majority of the apex predators across much of South Africa.[2] In contrast, the mesopredators, black-backed jackal (*Canis mesomelas*) and caracal (*Caracal caracal*) are by all accounts thriving, at least as measured by their impact on livestock production. Increasingly, attempts to manage livestock predation give rise to deep polarisations, particularly between animal rightists and livestock producers, which further confounds an already complex situation. This complexity hampers the development of policy and regulations with regard to managing livestock predation. A recent global review of the scientific merit of studies on the efficacy of various predator control interventions highlighted the paucity of adherence to acceptable scientific methods in these studies, and recommends 'suspending lethal control methods' while appropriately designed studies are undertaken.[3] Treves et al.[3] did not identify any valid (by their criteria) studies undertaken in South Africa. This example highlights the need for a scientifically robust basis for policy and management of livestock predation issues.

We support the principle of evidence-based policy and management, and propose that a formal scientific assessment[4,5] will provide scientifically robust and policy-relevant insights to address this challenge. Here we provide the framework for such an assessment on livestock predation in South Africa, and anticipate some of the emergent values of this assessment.

The Nelson Mandela University, through the Centre for African Conservation Ecology, has partnered with the Department of Environmental Affairs; the Department of Agriculture, Forestry and Fisheries (through the Red Meat Research Development Planning Committee); Cape Wools; and the SA Mohair Growers Association, and initiated the process of undertaking a scientific assessment on the issue of predation on livestock in South Africa (hereafter PredSA). PredSA was formally launched in June 2016 when it received the endorsement of the Minister of Environmental Affairs and the Department of Agriculture, Forestry and Fisheries. Approximately ZAR2.5 million has been committed to the assessment, which is anticipated to be 18 to 24 months in duration, starting May 2016.

PredSA will be conducted as an independent, science-based assessment, along the lines of the Elephant Management Scientific Assessment.[6] The assessment process will be grounded in five driving principles: *legitimacy*, *saliency* and *credibility*, which are underpinned by *transparency* and which is broadly *participatory.*

An independent six person Process Custodian Group has been appointed with the sole function of ensuring that the process of conducting the assessment is fair. The lead authors have been identified and the first workshop, which deals with the scoping and structure of the assessment, has been conducted. The next step, that of crafting an initial First Order Draft, is underway with the full complement of authors anticipated to be about 50 individuals.

PredSA will be compiled by recognised experts from academia and management who volunteer their input. The coverage will be comprehensive and include diverse topics in order to provide the context and detail that are relevant to policy development.

Understanding an issue requires a historical perspective, and thus the historical background to the longer-term predator–livestock interactions – contextualising historic socio-political and economic changes – within what is now the Republic of South Africa will be addressed. From the pre-colonial era onwards, human activities – specifically around pastoralism – have been negatively impacted by predation on domestic livestock with conflict as the usual consequence. This long-term perspective will also highlight how views, policy and approaches to livestock predation have changed.

Knowing the role players is key to managing them. Black-backed jackal and caracal are the dominant predators of livestock in South Africa.[1] Thus, PredSA will, in particular, explore the specific biological and ecological aspects of these two species that determine their role as livestock predators. A cornucopia of other species is implicated in livestock predation in South Africa, including lion (*Panthera leo*), leopard (*Panthera pardus*), cheetah (*Acinonyx jubatus)*, Cape fox (*Vulpes chama*), African wild dog (*Lycaon pictus*), side-striped jackal (*Canis adustus*), spotted hyena (*Crocuta crocuta*), brown hyena (*Hyaena brunnea*), serval (*Leptailurus serval*), baboon (*Papio ursinus*), honey badger (*Mellivora capensis)*, bushpig (*Potamochoerus larvatus*), crocodile (*Crocodylus niloticus*), feral domestic dogs (*Canis lupus familiaris*), and various corvids and raptors. For all these species (and any others that may be identified through PredSA processes), we will evaluate the evidence of them attacking livestock (excluding poultry), identify which livestock are attacked, and categorise the severity of this predation. The ecology and behaviour of the main livestock predators will be reviewed to determine how these affect the interaction with livestock. PredSA will allow us to determine which factors play a role in livestock predation, as well as to identify any potential gaps in the knowledge base which require future research.

What predation control methods are available and what are their outcomes and desirability? Historically, and currently, predation management has focused on ways to remove (so-called lethal control) or exclude the problem species from a specific area.[7] However, emerging evidence suggests that not all predators are problem animals, and that territorial individuals may act as a catalyst to exclude potential problem individuals.[8] Public opinion against lethal control has grown[9], while new insights have been gained into the environmental or ecological effects of such control (for example see Minnie et al.[10]). As a result, there seems to be a shift from attempts to eradicate predators to non-lethal methods to reduce predation, and to approaches in which only problem individuals are removed. These interventions, their efficacy and trends in their application need to be analysed and presented in a policy-relevant framework.

South Africa is not alone in experiencing problems with predation of livestock, as this phenomenon emerges across the world wherever livestock and predators co-occur. It is therefore fitting to identify and assess the various management strategies and internationally recognised best practices employed beyond our borders, and identify those that are most likely to be effective under South African circumstances. Special attention will be paid to those studies which replicate our semi-arid conditions and the types and sizes of predators involved, e.g. comparing coyotes (*Canis latrans*) and dingoes (*Canis lupus dingo*) to jackals. Given that South Africa lags behind the rest of the world in terms of scientifically evaluating the efficacy of livestock predation interventions[3,11], these lessons should extend to the design and implementation of sound scientific experiments, so that the outcomes are credible and applicable.

Although black-backed jackal and caracal are heavily persecuted in South Africa, we lack a clear understanding of the ecosystem-level consequences this persecution may have. Both these species vary in their roles in food webs, ranging from mid-ranking mesopredators that regulate small mammal and rodent communities[12] to apex predators that impact on a wide range of prey[13]. PredSA will review the functional role of black-backed jackal and caracal across a range of landscapes, from those inclusive of apex predators to those dominated by humans and livestock. In the western USA, coyote persecution resulted in a reduction of available forage for cattle because of high levels of competition with abundant lagomorphs.[14] Therefore, understanding the functional role of mesopredators in agro-ecosystems provides a more holistic basis for management actions and predicting their outcomes.

The basis of conflict around livestock predation is the impact this predation has on human well-being, or perceptions thereof. The presence of carnivores on rangelands can lead to losses of stock, expenditure on measures to prevent these losses, or, depending on the level of investment in and/or efficacy of the latter, costs incurred through a combination of the former and latter. Both stock losses and investments in predator control measures translate into a reduced bottom line for farmers. This affects wildlife-based operations as well as small and large stock farmers. Economic theory suggests that predator control efforts would be expected to kick in once losses reach a certain level, but actual investment decisions are influenced by a range of social factors and perceptions. The consequences of predation mitigation interventions are not well understood, which results in the investments in these actions not necessarily delivering intended outcomes. Livestock predation is widespread, thus losses incurred by farmers are believed to have a significant impact on the economic value of the industry, which translates into loss of employment opportunities as well as income, and, for some, their livelihoods. Conversely, changes in biodiversity brought about by changes in the relative abundance of livestock predators may impact on producer and consumer surplus associated with rangeland-based activities, as well as on society in less tangible ways. There are distributional issues too. While predator-control activities can provide direct income-earning opportunities in rural areas, the well-being of other members of society may be negatively affected by these activities.

A review of the legal framework in terms of the law relevant to management and control of predators in South Africa, and an analysis of its shortcomings that may impact on the efficacy of management practices and policy, is required. PredSA will outline the current status of South African law applicable to the management and control of predators, and assess the legislative gaps and contradictions in order to assist the relevant authorities in the development of policy and regulations. In making policy decisions, the regulatory authority is often confronted with differing interpretations of the law that appear to present options or alternative approaches. This review is intended to assist policymakers to develop legislative mechanisms that are in accordance with the law, or, when the law is seen to be lacking, to provide a sound legal basis to implement policy or legislation that is aligned with the constitution and legislation. The conclusions and recommendations will be drawn from legislation as it is generally accepted to be, and on interpretations of common law, as well as a consideration of customary law.

From an ethical perspective, the key issue with respect to livestock and predation is that it entails conflict. There are obvious conflicts of interest between livestock owners and predators. Furthermore, local communities, wildlife conservationists, ecotourists, and farmers have interests that differ and may clash. Conflicts of interest often lead to more worrying kinds of conflict, with those seeking to protect their interests ending up at loggerheads with one another. We also often find ourselves torn between competing moral obligations: our duties to our fellow humans may conflict with our duties to other species or the environment as a whole. This dilemma represents a significant challenge for policymakers. In such situations, the best that they can do is to try to carefully weigh up all of the ethical obligations and the competing interests, to come up with approaches that result in the best overall outcomes for all relevant stakeholders. This goal cannot be achieved without being in possession of the most relevant information required to be able to do this kind of weighing up. The better equipped decision-makers are with all of the relevant data, facts, perceptions, points of view and other relevant information, the better the policies they will be able to devise. In fact, it is an ethical obligation for policymakers to ensure that they have done their best to gather all of the necessary information to be able to make the most appropriate decisions. This is why this scientific assessment is not just important – it is also imperative.

In addition to bringing together the information and views relevant to livestock predation and its management in South Africa in a policy relevant fashion, this scientific assessment will deliver a number of further benefits. PredSA also provides an opportunity for those with conflicting views on predator management approaches to recognise and understand the alternative perspectives, and the broader implications of management approaches. This assessment process should therefore turn this area of tension into a commitment to finding a shared solution to the problem. This relaxation of tension is one of the outcomes of the scientific assessment on elephant management in South Africa.[6] Prior to this assessment, the so-called elephant debate was driven by strong views and tensions, whereas much of this acrimony has subsequently declined. The Norms and Standards for Elephant Management[15], developed in parallel with the assessment, have now been implemented.

Another emergent aspect of the PredSA assessment is that it will identify agreed-upon gaps in our knowledge. Such gaps may reflect specific hypotheses that require testing, or information that is required to test such hypotheses. Other gaps in our knowledge may relate more to the social dimensions of the issues related to predation, establishing a need for exploratory, qualitative research. These identified areas can be used to guide research needs and priorities in predation management – for both researchers and research funders. Given the multidisciplinarity of the assessment process, it can be predicted that novel and stimulating areas of research will be identified, and research synergies previously not thought of will be generated.

Adaptive management – the approach whereby management interventions are treated as experimental tests of predictions arising from hypotheses of complex systems' behaviour[16] – has the ability to advance the understanding of such systems and thereby assist managers to achieve desired goals. The management of livestock predation is ideally undertaken through such adaptive management approaches, as the system is complex and we have much to learn, including the validity of prevailing hypotheses or hypotheses emerging from PredSA. Thus, outcomes of PredSA will guide adaptive management approaches and strengthen relations between livestock managers and scientists, as the former can be seen as running a series of experimental manipulations which yield data for the latter to interpret. This relationship adds a further opportunity for the strengthening of research capacity in South Africa, where every livestock farmer may become a 'citizen scientist'.

We conclude that science can and must provide valid inputs into the challenges and policy needs of livestock predation management in South Africa through the PredSA scientific assessment process. Furthermore, we predict that PredSA will give a much needed boost to building transdisciplinary research capacity and raise the standards of research on livestock predation and management in South Africa.

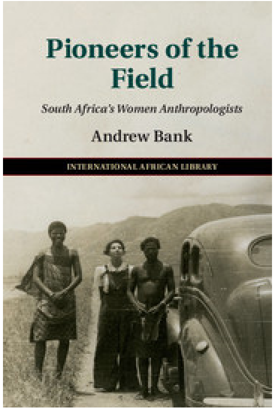
References

1. Van Niekerk HN. The cost of predation on small livestock in South Africa by medium sized predators [MSc thesis]. Bloemfontein: University of the Free State; 2010.
2. Boshoff AF, Landman M, Kerley GIH. Filling the gaps on the maps: Historical distribution patterns of some larger mammals in part of southern Africa. Trans R Soc S Afr. 2016;71:23–87. https://doi.org/10.1080/0035919X.2015.1084066
3. Treves A, Krofel M, McManus J. Predator control should not be a shot in the dark. Front Ecol Environ. 2016;14:380–388. https://doi.org/10.1002/fee.1312
4. Scholes RJ, Schreiner G, Snyman-van der Walt L. Scientific assessments: Matching the process to the problem. Bothalia. In press 2017.
5. Von der Heyden S, Lukey P, Celliers L, Prochazka K, Lombard AT. Science to policy – Reflections on the South African reality. S Afr J Sci. 2016;112(11/12), Art. #0183, 6 pages. https://doi.org/10.17159/sajs.2016/a0183
6. Scholes RJ, Mennell KG, editors. Elephant management: A scientific assessment for South Africa. Johannesburg: Wits University Press; 2008.
7. Du Plessis JJ. Towards a sustainable management strategy for *Canis mesomelas* and *Caracal caracal* on rangeland [PhD thesis]. Bloemfontein: University of the Free State; 2013.
8. Avenant NL, Du Plessis JJ. Sustainable small stock farming and ecosystem conservation in southern Africa: A role for small mammals? Mammalia. 2008;72:258–263. https://doi.org/10.1515/MAMM.2008.041
9. Natrass N, Conradie B. Jackal narratives: Predator control and contested ecologies in the Karoo, South Africa. J S Afr Stud. 2015;41:753–771. https://doi.org/10.1080/03057070.2015.1049484
10. Minnie L, Gaylard A, Kerley GIH. Compensatory life-history responses of a mesopredator may undermine carnivore management efforts. J Appl Ecol. 2016;53:379–387. https://doi.org/10.1111/1365-2664.12581
11. Coppinger R, Coppinger L, Langeloh G. A decade of use of livestock guarding dogs. Proc Vertebrate Pest Conf. 1988;13:209–214.
12. Avenant NL. The caracal, *Felis caracal* schreber, 1776, as predator in the West Coast National Park [MSc thesis]. Stellenbosch: University of Stellenbosch; 1993.
13. Klare U, Kamler JF, Stenkewitz U, Macdonald DW. Diet, prey selection, and predation impact of black-backed jackals in South Africa. J Wildl Manage. 2010;74:1030–1042. https://doi.org/10.2193/2009-211
14. Ranglack DH, Durham S, Du Toit JT. Competition on the range: Science vs. perception in a bison–cattle conflict in the western USA. J Appl Ecol. 2015;52:467–474. https://doi.org/10.1111/1365-2664.12386
15. South African Department of Environmental Affairs. National norms and standards for the management of elephants in South Africa. Government Gazette 30833; 2008.
16. Biggs HC, Slotow R, Scholes RJ, Carruthers J, Van Aarde R, Kerley GIH, et al. Towards integrated decision making for elephant management. In: Scholes RJ, Mennell KG, editors. Elephant management: A scientific assessment for South Africa. Johannesburg: Wits University Press; 2008. p. 537–586.

Let us now praise famous women: Revising South Africa's foundational anthropology narrative

BOOK TITLE:
Pioneers of the field: South Africa's women anthropologists

BOOK COVER:



AUTHOR:
Andrew Bank

ISBN:
9781107150492 (softcover)

PUBLISHER:
Cambridge University Press, Cambridge; GBP64.99

PUBLISHED:
2016

REVIEWER:
Karen Tranberg Hansen

AFFILIATION:
Department of Anthropology, Northwestern University, Evanston, Illinois, USA

EMAIL:
kth462@northwestern.edu

HOW TO CITE:
Tranberg Hansen K. Let us now praise famous women: Revising South Africa's foundational anthropology narrative. S Afr J Sci. 2017;113(3/4), Art. #a0197, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0197

Women played prominent roles in founding anthropology in South Africa, according to historian Andrew Bank, yet their contribution has been nearly written off in conventional accounts of the founding fathers of the discipline. Winifred Tucker Hoernlé, Monica Hunter Wilson, Ellen Hellmann, British anthropologist Audrey Richards during her years at Wits (1938–1940), Hilda Beemer Kuper before she emigrated (1961) and Eileen Jensen Krige were far more influential in establishing the discipline than has been recognised. With the exception of Kuper and Richards, these women remained in South Africa and developed the discipline from within, under adverse intellectual and political circumstances. Spanning a century (1985–1995), their lives and scholarship unfolded against the backdrop of the institutionalisation of apartheid, which has complicated their legacy. Viewed largely by later generations as functionalist in theoretical orientation and conservative in political outlook, their contribution to the study of changing South African society has been marginalised.

Bank revises and complicates this drab narrative, piecing together well-known and previously unavailable details about the personal and intellectual lives of a group of women trained or influenced deeply by Winifred Hoernlé, who held the first university position in anthropology in South Africa at Wits. Praising her unusual gift as a teacher, her scholarship on social and cultural change, her work on social causes and welfare, and her sociability and humanity, they present Hoernlé as the 'mother' of South African anthropology in their recollections. Instilling joy in field research, she sent them all to the London School of Economics to participate in Bronislaw Malinowski's fieldwork seminars. As the mentor and role model, Hoernlé turned their focus to urban and applied anthropology in a liberal and humanist vision of society and race relations.

For his account, Bank draws on an impressive body of sources, including the women's scholarship, publications and writing, some of it unpublished field notes and papers, from public and private sources, including correspondence, photographs, interviews and communications with relatives. The chapters identify background experiences that shaped personal outlooks and analyses published and unpublished writings, often revealing little-known sides. Ellen Hellmann left no personal papers at all, yet Bank weaves together an extraordinarily engaging account based on her public life and personal interviews – what is in fact my favourite chapter.

Building up the Anthropology Department at Wits, Winifred Hoernlé introduced the new theory of structural-functionalism in British social anthropology to South Africa through her collaboration with A.R. Radcliffe-Brown (1923–1925). She had conducted research among the Nama in German South West Africa before World War I, returning during the early 1920s to do pioneering urban fieldwork in the Windhoek location. She and Radcliffe-Brown worked actively together, so much so that Bank makes the case for co-production in some of Radcliffe-Brown's seminal essays. During her tenure, anthropology in South Africa was transformed as Hoernlé and her 'daughters' undertook field research for their investigations of cultural tradition and social change in urban as well as rural settings. Following her resignation in 1937, Hoernlé's second career made her a leading activist in social welfare programmes involving white, Indian and African women and she promoted women's rights and liberalism in organisations at national level.

Bank's focus on Monica Wilson's relationship with Hoernlé adds fresh perspectives to an anthology about Wilson and her interpreters, which he co-edited recently (2013). Wilson received her anthropology training at Cambridge. During her research in Pondoland (1931 and 1932), Hoernlé advised her about fieldwork and helped her find funding. As their relationship became close, Wilson became Hoernlé's 'intellectual daughter.' Bank foregrounds Christianity as the distinctive feature of Wilson's anthropological identity as well as a theme in much of her scholarship, noting her turn toward history in later work. His discussion of her three-decade long university career demonstrates her significant involvement in university politics as the first, and only, woman in the university senate at the University of Cape Town.

Ellen Hellmann is best known to anthropologists as author of the pioneering urban study of an African slum-yard in Johannesburg (Rooiyard) conducted in the 1930s but published only in 1948. Her Jewish background shaped her empathy for the marginalised and influenced her Jewish and anti-apartheid political activism. Bank's close reading of both her unpublished doctoral thesis on education and African township youth and the published report reveals an innovative field researcher, pioneering the documentation of youth culture. Her later achievements as a fieldworker and author of essays on urban issues and race relations, popular and scientific, have not been well recognised, perhaps because she followed an activist rather than a university-based career, associated for years with the South African Institute of Race Relations.

The biographical studies of Audrey Richards bypass her three years at Wits, which also go unmentioned in historical accounts of anthropology in South Africa. Showing her role in consolidating the liberal vision established by Hoernlé, Bank describes her developing an anthropology that was ethnographically inspired, practical and applied as well as theoretically oriented. In addition to getting students into the field, while at Wits she also wrote several essays based on her work on the Bemba of Northern Rhodesia. Emphasising intellectual friendships, Bank notes her promotion of the work of young women anthropologists, characterising her as the 'surrogate mother' of South African anthropology.

Hilda Kuper and her husband Leo left South Africa to teach at the University of California in Los Angeles in 1961 because of their left-liberal politics. After describing her early urban fieldwork in Johannesburg and historically oriented scholarship in Swaziland, Bank focuses on a lesser-known aspect of Kuper's work – her creative writing.

Showcasing her substantive ethnographic fiction and poetry, he offers a detailed appraisal of her changing fiction, its relation to fieldwork and her own background. He reminds us of the research she conducted in Indian communities in Durban in the late 1950s that rarely is mentioned in accounts of South African anthropology. Her Jewish identity shaped her sympathetic relations with Indian women. As one of the most prolific anthropologists of her generation, Kuper taught and mentored a new generation of women anthropologists, both in South Africa and the USA.

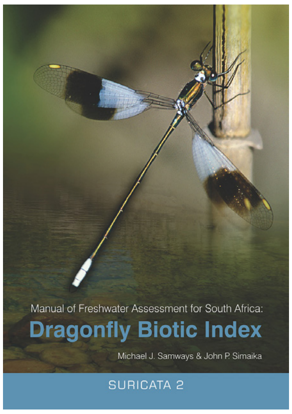
Because of her long career in anthropology at South African universities – although she retired in 1970 she remained active much longer – Eileen Jensen Krige is 'the university woman' in Bank's account. Trained as a teacher, she became Hoernlé's first student, undertook urban fieldwork in the 1930s, and, along with her husband Jack, she conducted fieldwork among the Lovedu, which resulted in a co-authored book. Bank remarks on her innovative methodology and her close relationship with Simeon Modjadji in half a century of collaborative field research. Like Hoernlé, Krige was very active in social welfare projects. Through writing about African women and social change, she left a legacy of feminising the discipline and promoting the careers of African, Indian and white women students.

In his conclusion, Bank agonises over anthropology's amnesia: why the contribution of these women has been written out of the male-dominant canon in the history of South African and British functionalist anthropology, their works viewed as 'tribal' and historically static, and their politics supportive of segregation and apartheid in South Africa. He argues the opposite: that their lives and works demonstrate a profoundly humanist endeavour against the backdrop of the institutionalisation of apartheid policies from the late 1940s onwards when the politics of racism and Afrikaner nationalism hardened. Representing the liberal-functional tradition in South Africa, their works dealt with the changing relationship between tradition and social change, thus taking account of history. Balancing personal lives and public careers, they might not all have published conventional fieldwork-based monographs and scientific articles in disciplinary journals, yet they pioneered urban field research and pursued applied and welfare projects involving diverse racial groups. Most of them were highly innovative in methodological terms. While not active in the feminist movement, they advanced South African feminism by their focus on women, promoted the status of women in male-dominated universities and public life, and pursued more applied work in and beyond the university than their male peers. Hoernlé inspired them to work on welfare and anti-apartheid campaigns in a shared passion for anthropology. While liberalism may translate differently in the North, South Africans are likely to weigh in on this important book from a variety of angles. And to be sure, anthropologists, historians and African Studies scholars in South Africa and elsewhere will find much to ponder in this well-written and meticulously researched study of six women who shaped South African anthropology in profound ways.

Dragonflies as indicators of aquatic ecosystem health

BOOK TITLE:
Manual of freshwater assessment for South Africa: Dragonfly Biotic Index

BOOK COVER:



AUTHORS:
Michael J. Samways and John P. Simaika

ISBN:
9781928224051 (softcover)

PUBLISHER:
South African National Biodiversity Institute, Pretoria; ZAR200

PUBLISHED:
2016

REVIEWER:
Ferdinand C. de Moor[1,2]

AFFILIATIONS:
[1]Albany Museum, Grahamstown, South Africa

[2]Department of Zoology and Entomology, Rhodes University, Grahamstown, South Africa

EMAIL:
f.demoor@ru.ac.za

HOW TO CITE:
de Moor FC. Dragonflies as indicators of aquatic ecosystem health. S Afr J Sci. 2017;113(3/4), Art. #a0199, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0199

Fresh waters are the most threatened ecosystems in the world. In South Africa, water abstraction, alteration of riverine ecosystems through dam building, inter-basin water transfer, introduction of alien aquatic organisms and organic and industrial pollution have left few freshwater ecosystems in a natural state. Because of their size and interesting behaviour, dragonflies and damselflies (Order: Odonata) are considered to be the most charismatic of all aquatic insects. They have attracted the attention of amateur societies and clubs in the Americas, Europe and Japan. In South Africa, the study of dragonflies has been well established scientifically through the works of pioneering entomologists such as Elliot Pinhey and Boris Balinski. Warwick and Michelle Tarboton have done much to transform this information on South African Odonata to a more popular media through the production of several guides and fold-out identification guides. Michael Samways and John Simaika now take this one step further in developing an index that uses the identification of Odonata to species level, to assess and monitor the health of aquatic ecosystems.

Of the 162 recorded species of Odonata in the book, about 20% are endemic to South Africa and these species require special conservation efforts to ensure their survival. There are also sufficient numbers of species that show strong preferences for narrowly defined aquatic and riparian biotopes, which allows for the identification and refined description of species-specific habitats. Odonata are considered to be good indicators of environmental health and water quality because all the species within this order are dependent on water for the development of their pre-adult stages (commonly known as nymphs, naiads or larvae). With a wealth of research information and good distribution records of these insects, the concept of using dragonflies as surrogate indicators of the conservation status of the immediate environment where these insects are recorded, was evaluated by the authors. This resulted in the development of the Dragonfly Biotic Index (DBI).

The emphasis of the DBI is on identification of adult dragonflies and damselflies to species level. At the species level, the range of climatic and environmental conditions can be accurately assessed and a tolerance value for each species can be defined. This assessment is not possible when dealing with genera or families which comprise large numbers of species, some with very different and non-overlapping ecological requirements.

To calculate the DBI, each species is given a value (between 0 and 9) based on scores ranging from 0 to 3 derived for each of three sub-indices. The first sub-index evaluates the geographical distribution range of a species. A rare, narrowly distributed endemic species would score the highest value of 3 and a widespread common species would score 0. The second sub-index is based on the threatened status of a species and the five red-list categories developed by the International Union for Conservation of Nature, which range from Critically Endangered to Least Concern status. Global or national evaluations are used to allocate a value from 0 to 3 for each species. The final sub-index relates to the sensitivity of a particular species to anthropogenic disturbance of its natural habitat, such as alien invasive riparian vegetation, water pollution or abstraction. Species found in pristine undisturbed sections of river are given a score of 3, whereas species tolerant of disturbance would get a score of 0 or 1. The DBI can be used for assessing environmental health and conservation status of aquatic ecosystems or for monitoring changes that occur over time.

The habitat condition scale – a metric derived from the DBI – is used to assess the status of the 'biotope diversity' in an ecosystem. To develop this metric, at least 20 sites in an ecoregion need to be comprehensively evaluated for DBI scores. The total DBI value per site (summed DBI values of all species) is plotted against the average DBI value per site (total DBI/number of species). This plot is then used to select the top 10% of sites that qualify for the highest biotope diversity classification. An example is given for the SE Coastal belt ecoregion in which DBI scores higher than 46 and average DBI scores greater than 5.4 rank as the highest biotope diversity category whereas DBI scores lower than 25 and average DBI scores less than 3.5 rank as the lowest.

The book is well laid out and the authors explain in detail how each species' DBI is calculated and how the combined values for all species recorded at a site are used to obtain a DBI site score. The equipment needed to undertake a DBI survey is listed, and the optimal conditions for when and how surveys should be undertaken are clearly explained. In more sub-tropical realms, a minimum of two visits on warm, humid days during the late spring to autumn should cover the majority of species at a site. These visits do not need to be evenly spaced in time and can take place within a month. It is noted, however, that in the cooler southern and southwestern Cape, a survey during late spring to early summer and another one in late summer should be undertaken as there are distinct early- and late-season species in that region. The book also includes a wealth of valuable information on each species in tables, which include species distribution in all 31 ecoregions, the national and global red-listed status for each species and all sub-indices and the final DBI score for each species. At the end of the book there is also an alphabetical checklist of all taxa in the 10 recorded Odonata families.

There is a detailed full-page description for each species, with large, clear distribution maps, DBI scores and red-list threat status, as well as colour photographs of all species indicating diagnostic identification features and line drawings highlighting features useful for identification. For each species, a common vernacular name and the detailed binomial taxonomic name (including name and date of authority originally describing the species) are given. It is a pity that the authors did not include a key or some guide to help identify and separate the different families and genera. I recommend that readers acquire a copy of one of the identification guides produced by Tarboton and Tarboton[1-4] or Samways[5] in order to identify the species.

The authors explain that the 31 ecoregions of South Africa used to record dragonfly distribution ranges are based on terrestrial rather than aquatic ecoregions. This is not entirely correct because the original publication[6] defines these ecoregions as 'Level 1 River Ecoregions' based on physiography, climate, geology and soils, potential natural vegetation, and hydrology. These 31 ecoregions also differ somewhat from the 35 bioregions[7] which are based mostly on terrestrial plant communities. A comparison of the DBI with the SASS5 (South African scoring system version 5) Biotic Index to explain that these indices can be interchangeably used is not warranted. The SASS Biotic Index was exclusively designed and refined for assessing water quality in rivers from a large database of water quality and macroinvertebrate distribution data, gathered throughout South Africa over many years. SASS is largely based on the original Chutter's Biotic Index[8] that used detailed species identification and abundance enumeration of taxa. This biotic index approach was simplified for SASS to require only family-level identification and relative abundance of aquatic invertebrates collected in a specific manner from selected biotopes in a river.[9] The SASS protocol enabled practitioners to perform a survey of water quality, based on the family identification of living invertebrates collected and evaluated directly in the field. The DBI actually complements the SASS Biotic Index because of the more exacting species identification and the more integrated aspect of evaluating the overall environmental quality and conservation status of the ecosystem under investigation.

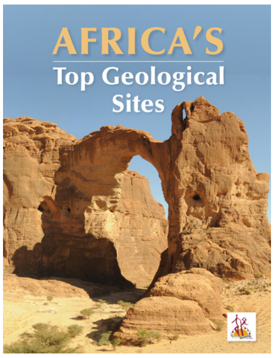
I would recommend this book to anyone interested in dragonflies and damselflies, as well as to anyone needing information on the assessment of the ecological status, environmental health, conservation and management of freshwater ecosystems.

References

1. Tarboton W, Tarboton M. A fieldguide to the dragonflies of South Africa. Modimolle: privately published; 2002.
2. Tarboton W, Tarboton M. A fieldguide to the damselflies of South Africa. Modimolle: privately published; 2005.
3. Tarboton W, Tarboton M. South African dragonflies: A quick guide. Johannesburg: Graphic Touch Guides; 2009.
4. Tarboton W, Tarboton M. A guide to the dragonflies & damselflies of South Africa. Cape Town: Struik Nature; 2015.
5. Samways MJ. Dragonflies and damselflies of South Africa. Sofia: Pensoft Publishers; 2008.
6. Kleynhans CJ, Thirion C, Moolman J. A level 1 ecoregion classification system for South Africa, Lesotho and Swaziland. Report no. N/0000/00/REQ0104. Pretoria: Resource Quality Services, Department of Water Affairs and Forestry; 2005.
7. Rutherford MC, Mucina L, Powrie W. Biomes and bioregions of southern Africa. Strelitzia. 2006;19:31–51.
8. Chutter FM. An empirical biotic index of the quality of water in South African streams and rivers. Water Res.1972;6:19–30. https://doi.org/10.1016/0043-1354(72)90170-4
9. Chutter FM. Report on the rapid biological assessment of water quality impacts in streams and rivers. Report no. 422/1/98. Pretoria: Water Research Commission; 1998.

BOOK TITLE:
Africa's top geological sites

BOOK COVER:



EDITORS:
C.R. Anhaeusser, M.J. Viljoen and R.P. Viljoen

ISBN:
978177584488 (softcover)

PUBLISHER:
Struik Nature, Cape Town; ZAR390

PUBLISHED:
2016

REVIEWER:
George Henry

EMAIL:
ghenry@uj.ac.za

AFFILIATION:
DST-NRF CIMERA, Department of Geology, University of Johannesburg, Johannesburg, South Africa

HOW TO CITE:
Henry G. Showcasing Africa's wonderful and varied geoheritage. S Afr J Sci. 2017;113(3/4), Art. #a0196, 1 page. http://dx.doi.org/10.17159/sajs.2017/a0196

Showcasing Africa's wonderful and varied geoheritage

The prestigious 35th International Geological Congress (IGC) was held at the Cape Town International Convention Centre from 27 August to 4 September 2016. This event, last held in South Africa in Pretoria in 1929 (the 15th IGC), was a resounding success, and congratulations must be extended to the Council for Geoscience of South Africa and the Geological Society of South Africa (GSSA) who undertook the bulk of the organisation.

As part of the event, the Geoheritage Committee of the GSSA saw an ideal opportunity to showcase Africa's superb geological heritage by producing a commemorative volume. An IGC Publications Panel, under the editorship of Professors Carl Anhaeusser, Morris Viljoen and Richard Viljoen, was constituted to canvass for articles from geoscientists who are experts in their fields with regard to the geoheritage of Africa. The editors are all noted geologists, now heading towards retirement, who have contributed much to the understanding of the geological history of South Africa, especially that of the Archaean period (greater than 2500 million years ago). They have made great strides in studying some of the oldest rocks in South Africa in the Barberton Mountain Land, as summarised by Professor Anhaeusser in Chapter 9 of the book.

The result of this collaborative effort is the publication under review. And it is overall a very good publication. A total of 41 authors contributed to 44 chapters in a full-colour, 312-page book that is profusely illustrated with photographs, maps and diagrams. It is impossible, as Professor Richard Viljoen states in his preface, to give an account of all of Africa's geoheritage in a single publication, but this is a mouth-watering sampling that leaves one hungering for Volume 2. It is not a publication to read from cover to cover (as there is too much information to digest in one go), but one in which to dip in and savour when one has the wanderlust and inclination to do so.

The book is more about Africa's geoheritage than it is about particular geological sites – despite the catchy title – which is exemplified by Chapter 13: 'Glossopteris in South Africa' by Dr Rose Prevec. Most of the chapters have concise geological and geomorphological descriptions of the geoheritage topics, and some have too-brief archaeological, historical and ecological paragraphs. Nevertheless, there is a bibliography for those who wish to delve further into the subject. As announced on a red and white sticker on my copy, 25 southern Africa sites are included, leaving many others elsewhere in Africa to be documented. As I am busy watching 'Voetspore' on television – an episode in which the adventure team is in Madagascar – it strikes me as a pity that the Publication Panel could not find a suitable author to document some of the island's varied geological make-up. The tsingy limestone terrain in the Bemaraha National Park would surely have made the list.

The style of writing is aimed at the informed layperson with an interest in things geological. Nonetheless, each article has undergone the strict academic review process to ensure scientific integrity. Some sections are not easy to read, but there is a glossary included that should assist overall comprehension. As a geologist myself, I understand that writing jargon-free popular articles is not easy, so I commend the authors on their effort, although some succeeded better than others.

The main criticism I have with the book is the question of scale. As a geologist, it is second nature to include a suitable scale bar – usually a standard geological hammer (about 33 cm in length), in all our photographs of rocks. But when I shared the photographs of the gold artefacts on Page 26 and was asked the size of the famous golden rhinoceros, I couldn't answer! Consequently, I was on alert for similar instances of a lack of a scale, and am sad to report that I found quite a few that need redressing in a second edition. Space precludes me from listing them all. The photograph of the fascinating orbicular granite at Diana's Pool in Zimbabwe has no indication of the size of the outcrop, but fortunately I know because my third-year Wits Geology class visited the site in 1975. Ah, the nostalgia!

Other niggling things include minor errors such as 'prophyries' for 'porphyries' on Page 28, 'Lembombo' on Page 52, and the disappearance of Rwanda and Burundi in Figure 33.1 on Page 226. These are not serious, unlike the edition of the Bible in 1631 that omitted the word 'not' from the seventh commandment! I hope that not many more will surface when I continue my perusal of the book and am overcome by the urge to fire up my temperamental 15-year-old 4x4 vehicle, affectionately named Daisy Disco, and hit the road…

I thoroughly recommend the book to all who have an interest in our continent's rich geological heritage. And hope that it will inspire visits to the more accessible sites as well as the more difficult to access ones. This book will certainly be kept in Daisy's glove compartment with my travel books for when I get the chance to do some leisurely geotourism.

The Acacia Gold Mining Company is to be thanked for the sponsorship of the publication, which has significantly reduced its price. Finally, the IGC Publications Panel and the production team at Struik Nature are to be congratulated on putting together a publication that deserves to be a bestseller.

Mycotoxigenic *Fusarium* species associated with grain crops in South Africa – A review

AUTHORS:
Ilze Beukes[1]
Lindy J. Rose[1]
Gordon S. Shephard[2]
Bradley C. Flett[3]
Altus Viljoen[1]

AFFILIATIONS:
[1]Department of Plant Pathology, Stellenbosch University, Stellenbosch, South Africa
[2]Institute of Biomedical and Microbial Biotechnology, Cape Peninsula University of Technology, Cape Town, South Africa
[3]Grain Crops Institute, Agricultural Research Council, Potchefstroom, South Africa

CORRESPONDENCE TO:
Lindy Rose

EMAIL:
lindym@sun.ac.za

DATES:
Received: 20 Apr. 2016
Revised: 23 Sep. 2016
Accepted: 30 Sep. 2016

KEYWORDS:
cereals; food safety; fungal contamination; mycotoxins; phytopathology

HOW TO CITE:
Beukes I, Rose LJ, Shephard GS, Flett BC, Viljoen A. Mycotoxigenic *Fusarium* species associated with grain crops in South Africa – A review. S Afr J Sci. 2017;113(3/4), Art. #2016-0121, 12 pages. http://dx.doi.org/10.17159/sajs.2017/20160121

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa); The Maize Trust; Winter Cereal Trust

Cereal grains include some of the most important crops grown in South Africa and play a major role in the local economy. Maize, wheat and sorghum are extensively consumed by humans and farm animals, and are also utilised in industrial processes. Grain crops that are grown commercially contribute up to 33% of the country's total gross agricultural production, whereas subsistence farmers grow grains mainly to sustain their families. In rural communities an average intake of maize grain of more than 300 g dry weight per person per day is not uncommon. The production of grains is often constrained by pests and diseases that may reduce their yields and quality. In South Africa, 33 mycotoxin-producing *Fusarium* species have been associated with grain crops. Mycotoxins, such as fumonisins and deoxynivalenol, have been found in levels exceeding the maximum levels imposed by the US Food and Drug Administration and the European Union and therefore pose a serious public health concern. We provide an extensive overview of mycotoxigenic *Fusarium* species associated with grain crops in South Africa, with particular reference to maize, wheat and sorghum.

Significance:

- Mycotoxigenic *Fusarium* species negatively affect the most important staple food crops grown in South Africa.
- Mycotoxin contamination has a direct impact on food safety and security.
- The genus *Fusarium* includes some of the most important mycotoxin-producing species.

Introduction

Grain crops grown in South Africa contribute between 25% and 33% of South Africa's total gross agricultural production.[1,2] The most commonly cultivated grain crops include maize (*Zea mays* L.), wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare* L.), sorghum (*Sorghum bicolor* L.), oats (*Avena sativa* L.), millet (*Pennisetum glaucum* L.) and rye (*Secale cereal* L.). Of these, maize is considered the most important and wheat the second most important.[2,3] The grains are utilised for food and livestock feed and, to a lesser extent, for malting purposes and bioethanol production.[1,2] Grains constitute the major portion of the total calorie intake of South Africans, across all age groups. The average consumption of maize by people older than 10 years varies from 762 g to 848 g cooked weight per person per day.[4]

The production of grain crops in South Africa is constrained by various abiotic and biotic stresses. Drought, nutrient deficiency, insect damage and diseases all cause a reduction in yield and grain quality.[5,6] One of the more important biotic stresses affecting maize, wheat and sorghum grain in the country involves the fungal genus *Fusarium*. The *Fusarium* sp. most commonly associated with these three grain crops is *F. graminearum* sensu lato (s.l.) Schwabe, also referred to as the *Fusarium graminearum* species complex.[7-10] Other *Fusarium* species affecting maize grain in South Africa include *F. verticillioides* (Sacc.) Nirenberg (syn. *F. moniliforme* Sheldon) and *F. subglutinans* (Wollenweber & Reinking) Nelson, Toussoun & Marasas, with *F. proliferatum* (Matsushima) Nirenberg occurring less frequently.[8,11,12] *Fusarium verticillioides* is also associated with grain mould of sorghum in South Africa.[9,13] Additional *Fusarium* species associated with sorghum include *F. thapsinum* Klittich, Leslie, Nelson & Marasas; *F. andiyazi* Marasas, Rheeder, Lamprecht, Zeller & Leslie; *F. nygamai* Burgess & Trimboli; and *F. pseudonygamai* O'Donnell & Nirenberg.[13] Fusarium head blight of wheat is associated with several species including *F. culmorum* (W.G. Smith) Sacc., *F. cerealis* (Cooke) Sacc. (syn. *F. crookwellense* Burgess, Nelson & Toussoun) and *F. avenaceum* (Fries) Saccardo.[10]

Infection of grain by *Fusarium* spp. does not only result in reduced yield and grain quality, but could lead to food safety concerns. Most *Fusarium* species produce one or more toxic secondary metabolites, commonly known as mycotoxins, in the grain.[14] *F. graminearum* s.l. produces type B trichothecenes (TCT-B) such as deoxynivalenol (DON) and nivalenol (NIV). Another important group of mycotoxins, the fumonisins (FUM), are produced by several *Fusarium* species (Table 1). Both *F. verticillioides* and *F. proliferatum* have been associated with the production of fumonisins in maize and sorghum grains in the country.[8,12,13]

The discovery of fumonisins in South African maize grain by Bezuidenhout et al.[15] sparked a significant interest in *Fusarium*-associated mycotoxins in the country and also worldwide. The objective of the current review is to give an overview of the information available on mycotoxigenic *Fusarium* species associated with grain crops in South Africa. We furthermore provide an outline on the production of the three most important grain food crops in South Africa: maize, wheat and sorghum.

Table 1: Mycotoxigenic *Fusarium* species associated with South African grain crops

Species	South African grain host	References	Mycotoxins associated with fungal species	References
Fusarium acuminatum	Barley, oats, sorghum, wheat	25,140,141	BEA, DON, HT-2, MON, T-2	14,140–143
F. andiyazi	Sorghum	13,33	FUM	11
F. anthophilum	Rice	144	BEA, FUM, MON	45,145
F. avenaceum	Barley, oats, sorghum, wheat	7,10,52,146	BEA, FusaC, MON	14,147,148
F. brachygibbosum	Wheat	10	Unconfirmed	149
F. cerealis (syn: *F. crookwellense*)	Wheat	10	DON, NIV, Fx, ZEA	14,150
F. chlamydosporum	Amaranth, maize, sorghum, wheat	10,49,52,151	HT-2, MON, T-2	45,152
F. culmorum	Barley, wheat	7,10	AcDON, DON, Fx, MON, NIV, T-2, ZEA	14,148,153–155
F. dimerum	Maize	49		
F. fujikuroi	Wheat	10	BEA, FUM, MON	11,26,156
F. globosum	Maize	157	BEA, FUM	158,159
F. incarnatum-equiseti species complex	Amaranth, maize, sorghum, wheat	10,48,52,151	BEA, DON, MON, NIV, ZEA	14,45,140,143,160
F. merismoides	Sorghum	52	ENN	161
F. napiforme	Millet, sorghum	32	FUM, MON	156,162
F. nygamai	Millet, sorghum	13,53	BEA, FUM, MON	45,156,163
F. oxysporum	Barley, maize, sorghum, wheat	10,140,164	BEA, FA, FUM, MON, ZEA	142,165,166
F. poae	Barley, maize, wheat	7,10,35,49	BEA, Fx, HT-2, NIV, T-2	14,27,45,148
F. proliferatum	Maize	12	BEA, FUM, MON	167–169
F. pseudograminearum	Ryegrass, wheat	10,103	AcDON, DON, Fx, NIV, ZEA	170,171
F. pseudonygamai	Sorghum	13	FUM, MON	11,13,172
F. semitectum	Sorghum	52	BEA, DON, MON, NIV, ZEA	46,140,173
F. solani species complex	Maize, sorghum, wheat	10,49,52	DON, FUM, T-2, ZEA	47,51 (unconfirmed)
F. subglutinans	Maize, sorghum	8,52	BEA, FA, FUM, MON	14,173–177
F. thapsinum	Sorghum	13	FA, FUM, MON	178,179
F. temperatum	Maize	180	BEA, FUM, MON	181
F. tricinctum species complex	Wheat	10	BEA, T-2, ENN, MON	14,27,148,182
F. verticillioides (syn: *F. moniliforme*)	Maize, rice, sorghum	8,13,146	BEA, FusaC, FUM, MON	80,183,184
F. graminearum species complex:	Amaranth	151	AcDON,DON, Fx, NIV, T-2, ZEA	45,50
F. acaciae-mearnsii	Wheat, sorghum	7, 9	3-ADON, NIV	185
F. boothii	Barley, maize, wheat	7	15-ADON	185
F. brasilicum	Wheat	7	3-ADON, NIV	185
F. cortaderiae	Wheat, sorghum	7, 9	3-ADON, NIV	185
F. graminearum	Barley, maize (roots), wheat	7,21	3-ADON, 15-ADON, NIV	185
F. meridionale	Maize (roots), sorghum, wheat	7,9,21	NIV	185

BEA, beauvericin; DON, deoxynivalenol; HT-2, HT-2 toxin; MON, moniliformin; T-2, T-2 toxin; FUM, fumonisins; FusaC, fusarin C; NIV, nivalenol; Fx, fusarenon-X; ZEA, zearalenone; AcDON, acetyldeoxynivalenol; ENN, enniatins; FA, fusaric acid; 15-ADON, 15-acetyldeoxynivalenol; 3-ADON, 3-acetyldeoxynivalenol

Grain crops in South Africa

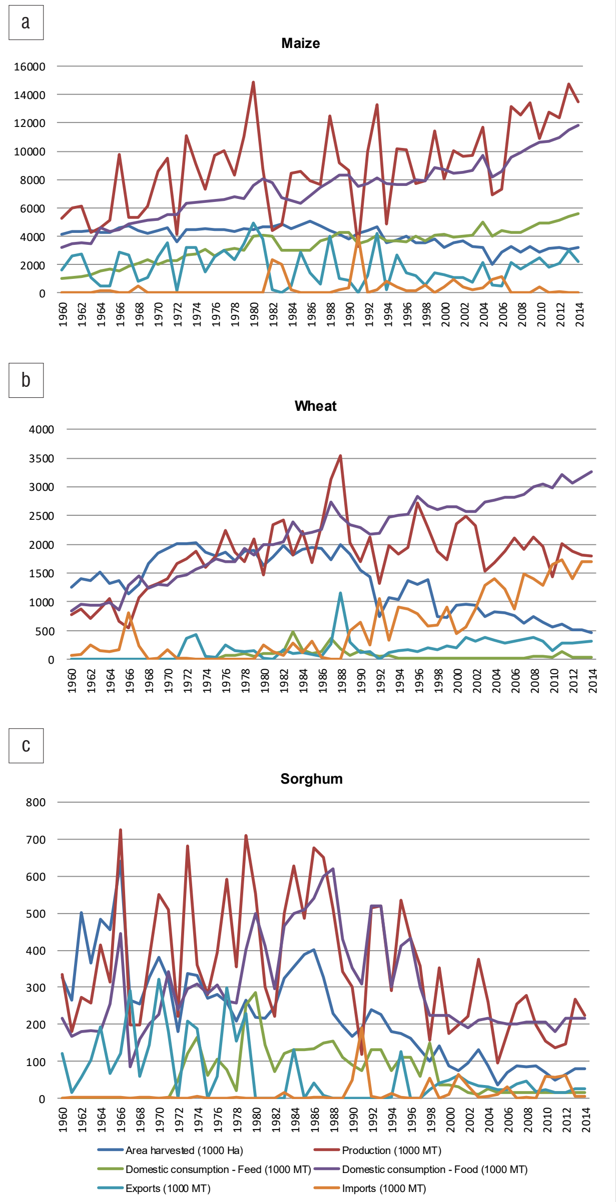
Maize

Maize forms the main staple food for the majority of South Africans, and constitutes a major component of animal feed. In 2014/2015, approximately 56% of the total area under maize cultivation (2 656 500 ha) comprised white maize, mainly used for human consumption, and 44% yellow maize, mostly used for animal feed.[2] The maize industry, therefore, is an important contributor to the economy of South Africa, both as an employer and generator of income.[1] In addition to its use as food and feed, maize is utilised in the manufacturing of paper, paint, textiles, adhesives, biodiesel, medicine and food.

Advances in maize cultivation practices – such as improved cultivars, effective crop rotation and enhancements in fertilisation and pesticide programmes – have steadily improved the yield per hectare. Whereas the total area harvested in South Africa has decreased from 4 118 000 ha in 1960 to 2 656 500 ha in 2014, yield has increased by 8 225 000 metric tons (Figure 1a).[2,16] The increase in production has ensured that the importation of maize has been minimised, and any surplus can be exported (Figure 1a), thus contributing towards generating foreign currency. The Free State (43%), North West (20%) and Mpumalanga (19.5%) Provinces of South Africa were the main production areas during the 2013/2014 production season for total white and yellow maize harvested.[2] Maize in South Africa is cultivated during the summer months with ideal planting times in November and December.

Maize production systems in South Africa can vary from resource-poor subsistence farming to small-scale and intensive commercial farming.[12,17] Chambers and Ghildyal[18] defined a resource-poor farm family as 'one whose resources of land, water, labour and capital do not permit a decent and secure family livelihood'. The Merriam-Webster online dictionary[19] defines subsistence farming as: 'Farming or a system of farming that provides all or almost all the goods required by the farm family usually without any significant surplus for sale.' The average yield per hectare recorded from 2008 to 2012 for non-commercial farmers was a meagre 1.3 tons/ha, while commercial farmers produced an average of 4.6 tons/ha.[2]

Diseases caused by fungal pathogens – aggravated by the use of inferior seed, monoculture and retaining crop residues – lead to reduced yields and lower grain quality.[12] A survey by Ncube et al.[12] during two production seasons determined that *F. verticillioides* was the most common *Fusarium* species associated with maize grain produced in a subsistence farming system, followed by *F. subglutinans* and *F. proliferatum*. Maize grain infected with these species was also contaminated with FUM, often at levels much higher than the maximum levels set by the US Food and Drug Administration (FDA) and the European Union (EU). A maize crop quality survey of commercially produced maize is performed annually by the Southern African Grain Laboratory (SAGL) with the financial support of The Maize Trust. Despite the general good quality of commercial maize, high levels of some mycotoxins can be found when weather or other conditions are favourable for fungal infection. Subsequently, mycotoxin contamination levels, in excess of the maximum levels allowed by the EU for maize intended for direct human consumption, have been found in commercially produced maize.[8,17,20] A 2-year survey of two susceptible maize cultivars, collected at 14 localities across South Africa, found a maximum total FUM level of 16 717 μg/kg, with an average of 2542 μg/kg and DON levels as high as 4731 μg/kg (average of 1031 μg/kg). Beauvericin (BEA) was recorded at a maximum level of 1507 μg/kg (average of 506 μg/kg) and moniliformin (MON) at a maximum of 1530 μg/kg and an average of 604 μg/kg.[8] Zearalenone (ZEA) has also been sporadically detected in South African maize. During the 2011/2012 season, only two samples analysed by SAGL tested positive. However, with an average of 249 μg/kg, they exceeded the maximum level of 100 μg/kg allowed by the EU.[20] The occurrence of these mycotoxins could be attributed to the presence of *F. verticillioides*, *F. graminearum* s.l. and *F. subglutinans*. The presence of high mycotoxin levels in commercial maize could possibly be attributed to the fact that commercial farmers still consider yield, not disease resistance, the number one criteria when deciding on a hybrid to plant. *Fusarium* spp. do not only cause ear rot, but can furthermore cause root, crown and stalk rot of maize, thereby causing additional yield losses.[21]



Values used to generate these graphs were obtained from: GrainSA[2] and the US Department of Agriculture[16]

Figure 1: Cultivation, usage and trade in South Africa between 1960 and 2014 of (a) maize, (b) wheat and (c) sorghum.

Wheat

Wheat is the second most important grain crop produced in South Africa and is also regarded as an important staple food. It serves as the second main provider of energy in the national diet after maize meal, even though more money is spent on bread annually (ZAR6.7 billion in 2000) than on maize food products (ZAR6.2 billion in 2000).[22] The majority of wheat cultivated in South Africa is bread wheat, with minor quantities of durum wheat produced for the production of pasta. Wheat is primarily used for human consumption (bread, biscuits, breakfast cereals, rusks), while the balance is used as seed for re-planting. Poorer quality wheat is marketed as animal feed and other non-food industrial uses such as the production of alcohol for ethanol, absorbing agents for disposable diapers, adhesives and starch on coatings.[23] Approximately 3900 commercial wheat farmers provide job opportunities to almost 28 000 people.[24]

Wheat production in South Africa can be divided into two different cultivation systems, each with their own adapted wheat varieties. In summer-rainfall areas, wheat is mostly cultivated under irrigation, and planted between mid-May to the end of July (Northern Cape, Free State, KwaZulu-Natal). In the Western Cape, a winter-rainfall area, wheat is mostly planted under dryland conditions between mid-April and mid-June. About 600 mm water per year is required for wheat cultivation and, in dry areas where zero tillage or minimum tillage are practised, stubble mulching is recommended for moisture conservation.[23]

The main wheat production areas in South Africa during the 2013/2014 season were the Western Cape (50%), Northern Cape (16%) and Free State (15%). Wheat is predominantly produced by commercial farmers with negligible amounts produced by small-scale and subsistence farmers, mainly because of high input costs and low yields, which results in smaller profit margins. Although the yield per hectare of wheat has shown a steady increase over the past 10 years (2.02 tons/ha in 2004 to 3.73 tons/ha in 2014)[2], the area harvested has decreased at a higher rate than the increase in yield could support, resulting in an overall reduction in production (Figure 1b). Lower production has led to an increase in the importation of wheat into South Africa to accommodate the drastic increase in domestic consumption (Figure 1b). Production of wheat in South Africa is constrained by several factors. Input costs have increased because of substantial increases in the cost of fertilisers and fuel, competitive international wheat prices and poor climatic conditions, amongst others.[23] Fertiliser costs in the Swartland wheat-producing area of the Western Cape can be as much as 30% of the total input cost and weeds may limit grain yields by approximately 20% annually.

Wheat is susceptible to a range of insect pests and diseases caused by plant pathogenic viruses, bacteria and fungi.[5-7,10] Several *Fusarium* species are associated with root rot, crown rot and head blight of wheat in South Africa, including *F. avenaceum*, *F. brachygibbosum* Padwick, *F. cerealis*, *F. chlamydosporum* Wollenweber & Reinking, *F. culmorum*, *F. graminearum* s.l., *F. incarnatum-equiseti* (syn. *F. equiseti* (Corda) Saccardo), *F. lunulosporum* Gerlach, *F. oxysporum* Schlechtendahl emend. Snyder & Hansen, *F. poae* (Peck) Wollenweber, *F. pseudograminearum* Aoki & O'Donnell, *F. solani* (Martius) Appel & Wollenweber emend. Snyder & Hansen and *F. tricinctum* (Corda) Saccardo.[7,10,25] The presence of some of these species may result in the contamination of the infected grain with mycotoxins such as DON, ZEA, BEA and MON.[14,26,27]

Sorghum

Sorghum is the fourth most important grain crop produced in South Africa after maize, wheat and barley, and the third most important food grain crop.[2,28] Barley is mostly used for malting purposes in the production of beer, and is not considered a major food crop in South Africa. Sorghum is indigenous to Africa and is considered a staple food in many rural communities in South Africa. Approximately 90% of commercially grown grain sorghum is used for human consumption in the form of beverages and food (e.g. malt and sorghum meal), while the remainder is used as animal feed.[28,29] Industrial uses of sorghum include wallboards, biodegradable packaging materials and the production of ethanol. The brewing industry is the main consumer of sorghum, and about 55% of the total domestic produce is used for the manufacturing of traditional African sorghum beer. Sorghum flour competes directly with maize meal as a breakfast cereal or as soured porridge, known as 'mabele'.[28] However, mabele has been found to have better nutritional value (9.7% protein, 1.6% fat) when compared to super maize meal (7.4% protein, 1.0% fat). In South Africa, sorghum cultivars are divided into three classes: Class GM includes sweet sorghum with a low tannin content, which is especially suitable for malting and milling purposes; Class GL includes sweet sorghum with a low tannin content, which is especially suitable for milling and animal feed purposes; and Class GH includes bitter sorghum with a high tannin content (bird resistant), which is used for industrial malting.[28]

The area under sorghum cultivation and the total production of sorghum in South Africa has been on a decline since 1986 (Figure 1c).[2,16] While maize and wheat increased in yield per hectare (Figure 1a,b), the same was not observed for sorghum (Figure 1c), for which the yield per hectare has remained mostly unchanged since 1995, fluctuating only with climatic changes.[2] This observation could be explained by the limited amount of research and development funds available to the sorghum industry, which is relatively small when compared to other major grain crops. A total of only 903 000 tons sorghum was produced in the 5 years from 2009 to 2014, in comparison to 8.9 million tons of wheat and 61 million tons of maize produced during the same period.[2]

Sorghum is planted from mid-October to mid-December in South Africa.[29] The Limpopo Province is the main sorghum-producing province, with limited production in other provinces such as Mpumalanga, North West, Northern Cape, Eastern Cape, KwaZulu-Natal and the Free State.[30] As with maize, sorghum farming systems vary from subsistence to intensive commercial farming, depending on farm sizes, production and marketing methods.[30] Subsistence farmers consume most of their products without measuring the area under production and yield. The average sorghum yield on smallholder farms is estimated to be 0.8 tons/ha[30], significantly lower than the 2.4 tons/ha produced on commercial farms[2]. The lower yield per hectare for subsistence farmers can be attributed to insufficient fertiliser and pest control programmes as well as soil cultivation and crop rotation practices, amongst others. These factors furthermore favour disease development by fungal pathogens, thus increasing the possibility of mycotoxin contamination.

Fusarium grain mould is a very important biological constraint to sorghum production worldwide, while Fusarium stalk and root rot may result in lodging, causing decreased yields.[13,31-33] Several mycotoxin-producing *Fusarium* species have been isolated from sorghum grain in South Africa. *F. andiyazi*, *F. nygamai*, *F. thapsinum* and *F. verticillioides*[13] are known FUM producers, while species within *F. graminearum* s.l. are TCT-B and ZEA producers[9].

The high consumption levels of up to 500 g/person/day[34] of inferior quality maize and sorghum by subsistence farmers pose a considerable threat to human health. Case studies have shown that the incidence of oesophageal cancer in areas where grain with high levels of FUM contamination is consumed is much higher than in other populations where FUM-contaminated food is not a staple.[35,36]

Mycotoxigenic *Fusarium* species affecting South African grains

The mycotoxin-producing *Fusarium* species first described from grain in South Africa was *F. culmorum*, which was isolated from the stems and roots of wheat grown near Stellenbosch, Western Cape, in the 1930s.[37] By the end of 1985, a total of 27 *Fusarium* species, either toxigenic or non-toxigenic, had been reported from a broad range of hosts in South Africa.[32,38] To date, 33 mycotoxigenic *Fusarium* species have been associated with local grain crops (Table 1). These species include *F. verticillioides, F. proliferatum* and *F. subglutinans*, which are commonly associated with Fusarium ear rot (FER) of maize, and *F. graminearum* s.l. that causes Gibberella ear rot of maize, Fusarium head blight (FHB) of wheat and barley and grain mould of sorghum (Table 1). Certain *Fusarium* species are associated with FER, FHB and Fusarium crown rot under specific climatic conditions. For instance, FHB of wheat is caused by *F. avenaceum*, *F. culmorum* and *F. poae* in the cooler regions, whereas *F. graminearum* is predominant in the warmer regions worldwide.[39] In South African maize, the FER pathogen *F. verticillioides* predominates in the warmer dry areas, while *F. subglutinans* is abundant in cooler areas. The Gibberella ear rot pathogen, *F. graminearum* s.l., is most prevalent in intermediate climate areas.[40] Mycotoxin-producing species such as *F. polyphialidicum* Marasas, Nelson, Toussoun & van Wyk and *F. sacchari* (E. J. Butler) W. Gams are known to occur on grain crops elsewhere in the world, but have, to date, not been found on South African grains. These two species have, however, been found in soil debris and sugarcane, respectively, in South Africa.[41,42]

FUM-producing *Fusarium* species, such as *F. verticillioides* and *F. proliferatum*, are often associated with maize and sorghum in South Africa (Table 1). Maize samples collected from 2001 until 2013 tested positive for FUM, sometimes at levels in excess of the maximum levels allowed by the EU.[20] More FUM and FUM-producing *Fusarium* species were found in maize grain produced commercially in warmer production areas of the Northern Cape, North West and Free State Provinces[17] than in the cooler production regions. Although FUM contamination of small grain cereals has been reported[43], this mycotoxin has not been found in wheat and barley in South Africa when employing a multi-mycotoxin screening method using ultra-performance liquid chromatography mass-mass spectrometry[44]. ZEA and TCT-Bs, however, have been found in both maize and wheat in the country,[20,44] but at higher levels and more frequently in maize than in wheat. The TCT-Bs and ZEA are primarily produced by *Fusarium* species within *F. graminearum* s.l.[45,46], and are commonly associated with Gibberella ear rot of maize, FHB of wheat and grain mould of sorghum[7-10].

The highly toxic mycotoxin, T-2 toxin, has until recently not been recorded in South Africa. T-2 toxin is most commonly produced by *F. sporotrichioides*, a fungus well-adapted to survive in colder countries.[47] Some T-2-producing *Fusarium* species, such as *F. poae* and *F. chlamydosporum* (Table 1), have periodically been isolated from wheat with FHB and maize with FER symptoms in South Africa.[35,48,49] The presence of T-2 toxin in local maize grain, recently reported by the SAGL[20], as well as its association with *F. verticillioides* and *F. graminearum*[50,51], requires further investigation.

Information on mycotoxin contamination of oats, sorghum and millet in South Africa is limited. Sorghum is affected by *Fusarium* species[32,33,52,53] that produce BEA, FUM, MON, TCT-B and ZEA, such as *F. avenaceum*, *F. chlamydosporum*, *F. nygamai* and the *F. solani* species complex (Table 1). Some of the same *Fusarium* species have also been associated with oats, millet or other less important grains such as amaranth. *F. verticillioides*, a common producer of FUM, also produces mycotoxins of lesser importance such as BEA, Fusarin C and MON (Table 1). The lower toxicity of these mycotoxins, and the relative complexity of multi-mycotoxin analysis,[54] limits the amount of data available on their occurrence in South African grains.

Role of mycotoxins in plant disease development

The role of mycotoxins in the interaction of fungi with plants is not always clearly understood. Some have, however, been shown to benefit the fungus.[55,56] The TCT-Bs, for instance, are phytotoxic and act as virulence factors on sensitive hosts, allowing the fungus to progress in plant tissue.[57] This effect was demonstrated by non-TCT-producing *F. graminearum* mutants that were pathogenic, yet caused less disease in maize than did wild-type TCT-producing isolates.[58,59] The virulence of *F. graminearum* and *F. culmorum* was also closely correlated with their DON and NIV deposition in wheat grain.[56] Adams and Hart[60], in contrast, reported that DON was not a virulence or pathogenicity factor for *F. graminearum* on maize, following virulence trials with non-toxic protoplast fusion *F. graminearum* strains.

FUM has been shown to be phytotoxic to maize seedlings, but its role in phytotoxicity, virulence and pathogenicity is unclear. The phytotoxicity of FUM was demonstrated by Williams et al.[61] and Arias et al.[62] who reported that FUM had a direct inhibitory effect on root growth, root hair development and other functions within the plant, whereas van Asch et al.[63] reported the mycotoxin to be phytotoxic to maize callus in culture. Symptoms were further induced when seedlings were watered with high concentrations of FUM in the absence of the pathogen.[62] Glenn et al.[64] demonstrated that FUM production by *F. verticillioides* is necessary for the development of foliar disease symptoms on maize seedlings. Desjardins et al.[65] acknowledged that FUM could play a role in virulence, but argued that it is not essential for pathogenicity to maize seedlings. These authors compared the offspring of a *fum*1^+ field strain of *F. verticillioides* with a high degree of virulence and that of a *fum*1^- field strain. They found that progeny with high levels of virulence were associated with FUM production, while highly virulent FUM-non-producing progeny were not observed. However, a highly virulent FUM-non-producing wild-type isolate was also identified, indicating FUM is not required for virulence. FUM non-producing mutants of *F. verticillioides*, generated by the disruption of the *FUM5* gene, have been as virulent on maize ears as their wild-type predecessor strains.[59]

Mycotoxins could also be involved in reproduction, fungal development and the colonisation of plant tissue. Disruption of a cyclin-like (C-type) gene, *FCC1*, resulted in reduced FUM B_1 synthesis and sporulation.[66] FUM is also believed to provide a competitive advantage to the fungus as it inhibits the mycelial growth of other fungal species in vitro.[67] The oestrogenic mycotoxin ZEA enhances perithecial production in *F. graminearum*, therefore the sexual development of the fungus is suppressed when ZEA synthesis is inhibited.[55]

Impact of mycotoxins on human and animal health

The mycotoxins most commonly found in South African grains include DON, FUM and ZEA.[8,12,20] DON, also known as vomitoxin because of its strong emetic effects after consumption, is one of the most widely distributed TCTs found in grain. When consumed by livestock, DON can lead to food refusal, vomiting, decreased weight gain and less effective feed utilisation.[68-70] These disorders then cause anorexia in pigs and other monogastric animals. Ruminants and poultry appear to be resistant to DON.[71] In humans, ingestion of DON-contaminated food has been associated with nausea, vomiting and diarrhoea.[72] Outbreaks of acute DON-associated gastrointestinal illness in humans have been reported in China in 1984/1985 and in India in 1987.[73] The ingestion of NIV, which is considered more toxigenic than DON, has resulted in decreased feed consumption, lower feed conversion efficacy and decreased liver weights when fed to chickens.[74,75] NIV and NIV-producing *F. graminearum* s.l. species have, however, been less frequently associated with South African grains.[7,8]

High levels of FUM in maize grain has been associated with leukoencephalomalacia. Leukoencephalomalacia is a fatal condition that causes the softening of brain tissue as a result of vascular insufficiency or degenerative changes in horses and rabbits.[76-78] FUM has also resulted in fatal pulmonary oedema in pigs and high tumour formation incidences in rats.[79-81] FUM was discovered following the association of *F. verticillioides*-contaminated maize grain with a high incidence of oesophageal cancer in the Transkei region (Eastern Cape Province) of South Africa.[35,80,82,83] Since then, the mycotoxin has also been associated with human oesophageal cancer in China and Italy and with prenatal birth defects and higher HIV transmission rates.[84,85]

ZEA is one of the most widely distributed *Fusarium* mycotoxins globally. Despite its relatively low acute toxicity, ZEA is biologically potent[86] and may cause reproductive disorders in farm animals[45,86,87]. ZEA-containing feed and fungal cultures fed to chickens and turkeys have resulted in significantly reduced egg production.[88,89] In humans, ZEA has been linked to hypoestrogenic syndromes and is believed to be an eliciting factor for advanced puberty development in girls.[90,91] The potential of ZEA to stimulate the growth of human breast cancer cells has also been demonstrated in vitro.[92]

The EU and FDA established maximum allowable levels for certain food contaminants, including mycotoxins, with the aim to reduce their presence in foodstuffs to the lowest levels reasonably achievable by means of good manufacturing or agricultural practices.[73,93] In addition to the USA and countries within the EU, more than 100 other countries have established mycotoxin regulations for at least aflatoxin B_1, mostly produced by *Aspergillus* spp., to aid in minimising food safety concerns.[94] Fewer countries regulate *Fusarium* mycotoxins, and in South Africa no restrictions for maximum allowable levels of any of the *Fusarium*-related mycotoxins in food and feed are governed by legislation.

Management of mycotoxigenic *Fusarium* species

Good Agricultural Practice is a collective set of international codes of practice which forms part of the Codex Code of General Principles on Food Hygiene.[95] These codes are concerned with all aspects of primary food production, including environmental protection and sustainability, economics, food safety, food quality and health security. It also complements the Hazard Analysis Critical Control Point food management system designed to limit food safety concerns, including food poisoning by mycotoxins.[95,96] The Good Agricultural Practice codes recommend practices for primary production of foodstuffs including fruits, vegetables, grains and legumes. Adherence to these codes of good practice does not only impact on food safety locally, but also influences international trade. Great attention should thus be given to these codes when deciding on an integrated disease management strategy to control *Fusarium* species and their associated mycotoxins in different grains produced in South Africa.

Pre-harvest control

Field preparation and cultivation practices play a central role in the management of *Fusarium* diseases and their associated mycotoxins.[95] The burial of residue plant material from a previous planting season by deep ploughing can reduce the primary inoculum that causes infections.[97] This is especially important when crops are affected by the same *Fusarium* species, such as *F. graminearum* s.l. on maize, wheat and sorghum grown in rotation. While minimum tillage has significantly decreased stalk rot and increased grain yield of sorghum in South Africa[31], it also has increased inoculum build-up of mycotoxigenic fungi in maize cropping systems[95]. Crop rotation with legumes, brassicas and potato could also significantly reduce *F. graminearum* s.l. levels.[98] Limiting plant stress to increase plant vigour by adhering to optimum plant dates, preventing drought stress and the optimal use of fertilisers have reduced *Fusarium* infection in a number of grain crops.[99-101] Control of alternative hosts for *Fusarium* species, which include grasses and weeds, can also reduce unwanted inoculum build-up.[95,102,103]

No fungicides are registered for the control of *Fusarium* grain diseases on maize, wheat or sorghum in South Africa.[104] Triazole fungicides such as metconazole and tebuconazole, however, have been shown to control FHB and DON contamination in wheat.[105] Control of mycotoxigenic *Fusarium* species in maize with fungicides, however, is difficult as ears are covered by tight husks which prevent contact with ear rot pathogens. Field trials in South Africa have reported no significant differences in the colonisation of maize grain by *F. verticillioides* or FUM contamination after application of protective fungicides such as the strobilurins, triazoles and benzimidazoles.[106] Chemical elicitors also failed to reduce FER and FUM contamination in maize.[107] As strict regulations on chemical pesticides and fungicide use are implemented to reduce human exposure and prevent environmental pollution, biological control has become more popular.[108] Non-pathogenic fungal antagonists such as *Phoma betae* A.B. Frank and *Trichoderma* spp*.* Persoon have reduced FHB and DON contamination under greenhouse conditions, but field results were variable and often failed.[99,109,110] In Ethiopia, 100% disease suppression of Fusarium root and crown rot of sorghum was reported after application of *Bacillus* spp. under greenhouse conditions.[111]

Disease resistance is the most proficient and environmentally safe management practice to reduce *Fusarium* diseases in grain crops. Several quantitative trait loci (QTLs) that underlie resistance to FHB have been mapped in wheat, and can be used for marker-assisted selection.[112-114] In South Africa, commercial wheat cultivars and breeding lines containing resistance QTLs derived from Sumai 3 lines with low levels of FHB and DON content were identified under field conditions.[115] The resistance of maize cultivars grown in South Africa to FER and FUM contamination are uncharacterised but resistant maize inbred lines were identified.[116] These inbred lines could be used as sources of resistance within maize breeding programmes. Mapping studies have previously identified QTLs associated with resistance to FER and FUM contamination in maize.[117,118] However, studies to identify possible QTLs for resistance to grain mould in sorghum were less frequent, but have shown some success.[119-121]

Unconventional methods to control plant diseases are becoming more common. Maize hybrids genetically modified with crystal (*Cry*) genes from the bacterium *Bacillus thuringiensis*, known as *Bt*-maize, reduced the feeding of stem borers and resulted in lower infection by *F. verticillioides* and *F. proliferatum* and subsequently reduced FUM contamination.[122] FUM detoxification has also been achieved by genetic modification of maize with a degradative enzyme originating from *Exophiala spinifera* and *Rhinocladiella atrovirens*.[123]

Post-harvest control

FUM and DON levels in grain do not increase significantly when grain is harvested at <14% moisture and when optimal moisture and temperature conditions and control of insect pests are maintained during storage.[124-126] The removal of mouldy, broken and underdeveloped kernels can also significantly lower mycotoxin levels in cereal grains. FUM levels of maize were reduced between 26.2% and 69.4% by sieving (<3 mm), and by 71% by separating mouldy from healthy kernels of maize produced by subsistence farmers in the former Transkei.[127,128] An 86% reduction of FUM was also achieved by the removal of *F. verticillioides*-contaminated maize kernels by flotation in water and sodium chloride, as these were less dense than sound kernels.[129] The separation of smaller, underdeveloped and shrivelled wheat kernels by means of the Carter dockage tester resulted in a 6–19% reduction of DON.[130]

Mycotoxins are mostly heat stable; however, the preparation of South African traditional maize porridge through normal household cooking can reduce FUM by 23%.[131] The washing of barley with distilled water has reduced DON levels by 69% and ZEA levels by 2%. In maize, DON levels were reduced by 65% and ZEA levels by 61%. A further reduction in DON and ZEA was achieved by using 1 M sodium carbonate solution for the first wash.[132] The conversion of mycotoxins into non-toxic products can also be achieved through physical or chemical processes. Chemical degradation of DON has been achieved by ammonia, calcium hydroxide, chlorine, hydrochloric acid, ozone, sodium bisulfite, and sodium hydroxide.[133-136] However, the large-scale application of these methods are hampered by costs, safety concerns and the negative impact on grain quality.[137] Biological detoxification, defined as the enzymatic degradation of mycotoxins or modification of their structure that leads to less toxic products, offers an alternative method to reduce the mycotoxin content in food and feed products.[138]

Discussion

Mycotoxigenic *Fusarium* species negatively affect the most important staple food crops grown in South Africa by reducing their yield and quality, and by contaminating the grain with harmful mycotoxins. These effects pose a serious threat to food safety and security for a rapidly expanding population. Efforts to manage harmful *Fusarium* species and their associated mycotoxins, both in commercial and subsistence farming systems, should therefore be made to sustain food production, to reduce health risks to humans and other animals, and to safeguard competitive international trade. A first step in achieving this aim could be the introduction of maximum levels for *Fusarium* mycotoxins in South African food and feed – a directive which has been conspicuously overlooked by the Departments of Health and Agriculture in the country.

A policy brief was compiled in 2009 to[139]:

> *assist national stakeholders in government and industry, as well as commercial and emerging farmers, in understanding and implementing a united monitoring programme for the prevention and control of mycotoxins in foods in South Africa.*

This brief recommended that DON and FUM be added to existing South African regulations in order to align with the guidelines adopted by most other mycotoxin-regulating countries.[94,139] Recently, South Africa has amended regulations regarding the tolerances for fungus-produced toxins in foodstuffs by limiting DON in grains to 2000 and 1000 μg/kg before and after processing, respectively. Maize grain, intended for further processing, is limited to 4000 μg/kg FUM while processed products, ready for human consumption, may not contain more than 2000 μg/kg of FUM.[186] Maximum levels for South Africa should be established by determining the general toxicity, haematotoxicity and immunotoxicity of the different mycotoxins as well as considering consumption levels of grain in the country. Incidences of mycotoxicoses, such as the outbreaks of DON-associated acute gastrointestinal illness in humans in China in 1984/1985 and in India in 1987[73], should also be taken into consideration. The biggest limiting factors in this undertaking would be the costs involved in an extensive regulation programme of foodstuffs, such as the laboratory analyses and the monitoring of revised mycotoxin legislation by health inspectors.

Health workers should be trained to identify symptoms exhibited by humans and animals in cases of mycotoxicoses. A serious call should be made on government to support mycotoxin research and to implement legislation on the levels of the different toxins present in foodstuff. The high intake of grains, in terms of both portion size and frequency, as staple foods by the majority of South Africans should be considered when determining allowable levels of contamination. Coordinated efforts should furthermore be made to launch public awareness campaigns through the distribution of educational information, in a responsible manner, without evoking public fear. These efforts should be particularly focused in subsistence farming communities, in which mycotoxicoses pose a genuine public health threat, as a high incidence of oesophageal cancer in the Transkei region of South Africa has been directly linked to high FUM contamination.[82]

Managing the incidence and severity of mycotoxin contamination in grains, to reduce human and animal health risks and to safeguard competitive international trade, requires continuous efforts to understand and subsequently control the *Fusarium* species responsible for the production of these mycotoxins. South Africa, with its internationally recognised track record in mycotoxin research, possesses the skills, expertise and motivation to continue to work towards food safety and security for all people.

Acknowledgements

Mycotoxin research at Stellenbosch University is funded by the Maize Trust, the Winter Cereal Trust and the Technology and Human Resources for Industry Programme (THRIP) of the National Research Foundation of South Africa. We declare no conflict of interest.

Authors' contributions

I.B. was responsible for the design and content of the work; for the collection, analysis and interpretation of the data; and for drafting the manuscript. A.V. was responsible for the conception and design of the work. L.J.R., G.S.S., B.C.F. and A.V. critically revised the manuscript. All authors approved the final version.

References

1. South African Department of Agriculture, Forestry and Fisheries. Pocket guide to South Africa 2011/12 [document on the Internet]. c2012 [cited 2015 Feb 24]. Available from: http://www.gcis.gov.za/sites/default/files/docs/resourcecentre/pocketguide/013_agriculture.pdf

2. GrainSA. Industry reports [homepage on the Internet]. c2015 [cited 2015 Feb 25]. Available from: http://www.grainsa.co.za/pages/industry-reports/production-reports

3. Goldblatt A. Agriculture: Facts and trends, South Africa [document on the Internet]. c2012 [cited 2015 Mar 30]. Available from: http://awsassets.wwf.org.za/downloads/facts_brochure_mockup_04_b.pdf

4. Nel JH, Steyn NP. Report on South African food consumption studies undertaken amongst different population groups (1983–2000): Average intakes of foods most commonly consumed. Pretoria: Directorate: Food Control, Department of Health; 2002. Available from: http://www.mrc.ac.za/chronic/foodstudy.pdf

5. Rybicki EP, Pietersen G. Plant virus disease problems in the developing world. Adv Virus Res. 1999;53:127–175. http://dx.doi.org/10.1016/S0065-3527(08)60346-2

6. Oerke EC. Crop losses to pests. J Agric Sci. 2006;144(1):31–43. http://dx.doi.org/10.1017/S0021859605005708

7. Boutigny A-L, Ward TJ, Van Coller GJ, Flett B, Lamprecht SC, O'Donnell K, et al. Analysis of the *Fusarium graminearum* species complex from wheat, barley and maize in South Africa provides evidence of specie-specific differences in host preference. Fungal Genet Biol. 2011;48(9):914–920. http://dx.doi.org/10.1016/j.fgb.2011.05.005

8. Boutigny A-L, Beukes I, Small I, Zuhlke S, Spiteller M, Janse Van Rensburg B, et al. Quantitative detection of *Fusarium* pathogens and their mycotoxins in South African maize. Plant Pathol. 2012;61(3):522–531. http://dx.doi.org/10.1111/j.1365-3059.2011.02544.x

9. Mavhunga M. *Fusarium graminearum* mycotoxins associated with grain mould of maize and sorghum in South Africa [MSc dissertation]. Bloemfontein: University of the Free State; 2013.

10. Van Coller GJ, Boutigny A-L, Rose L, Ward TJ, Lamprecht SC, Viljoen A. *Fusarium* spp. associated with head blight of wheat in South Africa. In: Acta Phytopathologica Sinica. 10th International Congress of Plant Pathology Abstracts; 2013 August 25–30; Beijing, China. Available from: http://www.isppweb.org/abstracts/interno_icpp_2013.pdf

11. Rheeder JP, Marasas WFO, Vismer HF. Production of fumonisin analogs by *Fusarium* species. Appl Environ Microbiol. 2002;68(5):2101–2105. http://dx.doi.org/10.1128/AEM.68.5.2101-2105.2002

12. Ncube E, Flett BC, Waalwijk C, Viljoen A. *Fusarium* spp. and levels of fumonisins in maize produced by subsistence farmers in South Africa. S Afr J Sci. 2011;107(1/2), Art. #367, 7 pages. http://dx.doi.org/10.4102/sajs.v107i1/2.367

13. Leslie JF, Zeller KA, Lamprecht SC, Rheeder JP, Marasas WFO. Toxicity, pathogenicity, and genetic differentiation of five species of *Fusarium* from sorghum and millet. Phytopathology. 2005;95(3):275–283. http://dx.doi.org/10.1094/PHYTO-95-0275

14. Bottalico A, Perrone G. Toxigenic *Fusarium* species and mycotoxins associated with head blight in small-grain cereals in Europe. Eur J Plant Pathol. 2002;108:611–624. http://dx.doi.org/10.1023/A:1020635214971

15. Bezuidenhout SC, Gelderblom WCA, Gorst-Allman CP, Horak RM, Marasas WFO, Spiteller G, et al. Structure elucidation of the fumonisins, mycotoxins from *Fusarium moniliforme*. J Chem Soc Chem Commun. 1988;11:743–745. http://dx.doi.org/10.1039/C39880000743

16. United States Department of Agriculture (USDA). Indexmundi [database on the Internet]. c2015. [cited 2015 Feb 20]. Available from: http://www.indexmundi.com/agriculture/?country=za&commodity=corn&graph=production

17. Janse van Rensburg B, McLaren NW, Schoeman A, Flett BC. Fumonisin producing *Fusarium* spp. and fumonisin contamination in commercial South African maize. Eur J Plant Pathol. 2015;141(3):491–504. http://dx.doi.org/10.1007/s10658-014-0558-7

18. Chambers R, Ghildyal BP. Agricultural research for resource-poor farmers: The Farmer-First-and-Last Model. Agric Admin. 1985;20:1–30. http://dx.doi.org/10.1016/0309-586X(85)90063-9

19. The Merriam-Webster dictionary [online]. Subsistence farming. [cited 2015 Mar 09]. Available from http://www.merriam-webster.com

20. South African Grain Laboratory. South African maize crop quality report 2012/2013 [document on the Internet]. c2013 [cited 2015 Mar 24]. Available from: http://www.sagl.co.za/Portals/0/Maize%20Crop%202012%202013/Maize%20Crop%20Quality%20Report%202012%202013.pdf

21. Lamprecht SC, Tewoldemedhin YT, Botha WJ, Calitz FJ. *Fusarium graminearum* species complex associated with maize crowns and roots in the KwaZulu-Natal province of South Africa. Plant Dis. 2011;95:1153–1158. http://dx.doi.org/10.1094/PDIS-02-11-0083

22. South African Department of Agriculture. Wheat [document on the Internet]. c2006 [cited 2015 Mar 10]. Available from: http://www.nda.agric.za/docs/FactSheet/Wheat06.pdf

23. South African Department of Agriculture, Forestry and Fisheries. Wheat production guideline [document on the Internet]. c2010 [cited 2015 Mar 11]. Available from: http://www.daff.gov.za/docs/brochures/prodguidewheat.pdf

24. National Chamber of Milling. Strategy document for the South African wheat to bread value chain [document on the Internet]. c2004 [cited 2015 Mar 10]. Available from: https://www.google.co.za/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiu7db3hfzRAhVJ7mMKHcnBCWQQFggYMAA&url=http%3A%2F%2Fwww.waterberg.gov.za%2Fdocs%2Fagriculture%2Fother%2FWHEAT%25252520Strategy%25252520(April%252525202004)%255B1%255D.doc&usg=AFQjCNFrgHBJi85HlJxF7Uw7gj5bamvxfA&sig2=DxyhgXuvmuaYMDhqFXkUqA&bvm=bv.146094739,d.d2s

25. Van Coller GJ, Sedeman Z, Lamprecht SC, Viljoen A. *Fusarium* species associated with crown rot of wheat in South Africa. In: 47th Congress of the Southern African Society for Plant Pathology; 2011 January 23–26; Kruger National Park, South Africa. Available from: https://www.researchgate.net/publication/269699542_Fusarium_species_associated_with_crown_rot_of_wheat_in_South_Africa

26. Moretti A, Logrieco A, Bottalico A, Ritieni A, Fogliano V, Randazzo G. Diversity in beauvericin and fusaproliferin production by different populations of *Gibberella fujikuroi* (*Fusarium* section Liseola). Sydowia. 1996;48:44–56.

27. Thrane U, Adler A, Clasen P-E, Galvano F, Langseth W, Lew H, et al. Diversity in metabolite production by *Fusarium langsethiae*, *Fusarium poae*, and *Fusarium sporotrichioides.* Int J Food Microbiol. 2004;95(3):257–266. http://dx.doi.org/10.1016/j.ijfoodmicro.2003.12.005

28. The Sorghum Section 7 Committee. Report on the investigation into the South African sorghum industry. A report by the Sorghum Section 7 Committee appointed by the National Agricultural Marketing Council [document on the Internet]. c2007 [cited 2015 Feb 25]. Available from: http://www.namc.co.za/upload/section_7_reports/SORGHUM%20S7%20FINAL%20REPORT%2003.07.mdi.pdf

29. South African Department of Agriculture. Sorghum production [document on the Internet] c2008. [cited 2015 March 11]. Available from: http://www.nda.agric.za/docs/Infopaks/FieldCrops_Sorghum.pdf

30. South African Department of Agriculture, Forestry and Fisheries. Sorghum production guideline [document on the Internet]. c2010 [cited 2015 Mar 10]. Available from: http://www.nda.agric.za/docs/Brochures/prodGuideSorghum.pdf

31. Flett BC. Crop rotation and tillage effects on yield and the incidence of root and stalk rot in sorghum (*Sorghum bicolor*). S Afr J Plant Soil. 1996;13:136–138.

32. Marasas WFO, Rabie CJ, Lübben A, Nelson PE, Toussoun TA, Van Wyk PS. *Fusarium napiforme*, a new species from millet and sorghum in southern Africa. Mycologia. 1987;79(6):910–914. http://dx.doi.org/10.2307/3807697

33. Marasas WFO, Rheeder JP, Lamprecht SC, Zeller KA, Leslie JF. *Fusarium andiyazi* sp. nov., a new species from sorghum. Mycologia. 2001;93(6):1203–1210. http://dx.doi.org/10.2307/3761681

34. Shephard GS. Impact of mycotoxins on human health in developing countries. Food Addit Contam. 2008;25(2):146–151. http://dx.doi.org/10.1080/02652030701567442

35. Marasas WFO, Wehner FC, Van Rensburg SJ, Van Schalkwyk DJ. Mycoflora of corn produced in human esophageal cancer areas in Transkei, southern Africa. Phytopathology. 1981;71:792–796. http://dx.doi.org/10.1094/Phyto-71-792

36. Jaskiewicz K, Marasas WFO, Van der Walt FE. Oesophageal and other main cancer patterns in four districts of Transkei, 1981-1984. S Afr Med J. 1987;72(1):27–30.

37. Doidge EM. The South African fungi and lichens to the end of 1945. Bothalia. 1950;5:1–1094.

38. Marasas WFO, Lamprecht SC, Van Wyk PS, Anelich RY. Bibliography of *Fusarium* (Fungi: Hyphomycetes) in South Africa, 1945-1985. Bothalia. 1985;1:97–104.

39. Doohan FM, Brennan J, Cooke BM. Influence of climatic factors on *Fusarium* species pathogenic to cereals. Eur J Plant Pathol. 2003;109(7):755–768. http://dx.doi.org/10.1023/A:1026090626994

40. Marasas WFO, Kriek NPJ, Wiggins VM, Steyn PS, Towers DK, Hastie TJ. Incidence, geographic distribution, and toxigenicity of *Fusarium* species in South African corn. Phytopathology. 1979;69:1181–1185. http://dx.doi.org/10.1094/Phyto-69-1181

41. Marasas WFO, Nelson PE, Toussoun TA, Van Wyk PS. *Fusarium polyphialidicum*, a new species from South Africa. Mycologia. 1986;78(4):678–682.

42. Leslie JF, Summerell BA, Bullock S, Doe FJ. Description of *Gibberella sacchari* and neotypification of its anamorph *Fusarium sacchari.* Mycologia. 2005;97(3):718–724. http://dx.doi.org/10.3852/mycologia.97.3.718

43. Busman M, Desjardins AE, Proctor RH. Analysis of fumonisin contamination and the presence of *Fusarium* in wheat with kernel black point disease in the United States. Food Addit Contam A. 2012;29(7):1092–1100. http://dx.doi.org/10.1080/19440049.2012.671787

44. South African Grain Laboratory. South African wheat crop quality report 2013/2014 [document on the Internet]. c2014 [cited 2015 Mar 24]. Available from: http://www.sagl.co.za/Wheat/Wheatreports/20132014Season.aspx

45. Logrieco A, Mule G, Moretti A, Bottalico A. Toxigenic *Fusarium* species and mycotoxins associated with maize ear rot in Europe. Eur J Plant Pathol. 2002;108(7):597–609. http://dx.doi.org/10.1007/978-94-010-0001-7_1

46. Marasas WFO, Nelson PE, Toussoun TA. Toxigenic *Fusarium* species: Identity and mycotoxicology. Philadelphia, PA: The Pennsylvania State University Press; 1984.

47. Leslie JF, Summerell BA. The *Fusarium* laboratory manual. Ames: Blackwell Publishing; 2006.

48. Marasas WFO. Mycotoxicological investigations on corn produced in esophageal cancer areas in Transkei. In: Pfeiffer CJ. Cancer of the esophagus. Boca Raton, FL: CRC Press; 1982. p. 29–40.

49. Chilaka CA. A survey of South African commercial feed grade maize for mycotoxins with particular reference to fumonisins using different analytical techniques [MTech thesis]. Johannesburg: University of Johannesburg; 2012.

50. Patience JF, Ensley SM, Hoyer S. Mycotoxin contamination of corn: What it is, what it does to pigs and what can be done about it. Iowa Pork Industry Center Fact Sheets Paper 18 [document on the Internet]. c2010 [cited 2016 Jan 23]. Available from: http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1017&context=ipic_factsheets

51. Matny ON. Screening of mycotoxin produced by *Fusarium verticillioides* and *F. proliferatum* in culture media. Asian J Agric Rural Dev. 2014;4(1):36–41.

52. Rabie CJ, Lübben A. The mycoflora of sorghum: *Sorghum caffrorum* malt. S Afr J Bot. 1984;3:251–255.

53. Marasas WFO, Rabie CJ, Lübben A, Nelson PE, Toussoun TA, Van Wyk PS. *Fusarium nygamai* from millet in southern Africa. Mycologia. 1988;80(2):263–266. http://dx.doi.org/10.2307/3807807

54. Herebian D, Zuhlke S, Lamshoft M, Spiteller M. Multi-mycotoxin analysis in complex biological matrices using LC-ESI/MS: Experimental study using triple stage quadrupole and LTQ-Orbitrap. J Sep Sci. 2009;32:939–948. http://dx.doi.org/10.1002/jssc.200800589

55. Wolf JC, Mirocha CJ. Regulation of sexual reproduction in *Gibberella zeae* (*Fusarium roseum* 'Graminearum') by F-2 (zearalenone). Can J Microbiol. 1973;19(6):725–734. http://dx.doi.org/10.1139/m73-117

56. Mesterhazy A. Role of deoxynivalenol in aggressiveness of *Fusarium graminearum* and *F. culmorum* and in resistance to Fusarium head blight. Eur J Plant Pathol. 2002;108(7):675–684. http://dx.doi.org/10.1023/A:1020631114063

57. Proctor RH, Hohn TM, McCormick SP. Reduced virulence of *Gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. Mol Plant Microbe Interact. 1995;8(4):593–601. http://dx.doi.org/10.1094/MPMI-8-0593

58. Harris LJ, Desjardins AE, Plattner RD, Nicholson P, Butler B, Young JC, et al. Possible role of trichothecene mycotoxins in virulence of *Fusarium graminearum* on maize. Plant Dis. 1999;83(10):954–960. http://dx.doi.org/10.1094/PDIS.1999.83.10.954

59. Proctor RH, Desjardins AE, McCormick SP, Plattner RD, Alexander NJ, Brown DW. Genetic analysis of the role of trichothecene and fumonisin mycotoxins in the virulence of *Fusarium*. Eur J Plant Pathol. 2002;108(7):691–698. http://dx.doi.org/10.1023/A:1020637832371

60. Adams GC, Hart LP. The role of deoxynivalenol and 15-acetyldeoxynivalenol in pathogenesis by *Gibberella zeae*, as elucidated through protoplast fusions between toxigenic and nontoxigenic strains. Phytopathology. 1989;79:404–408. http://dx.doi.org/10.1094/Phyto-79-404

61. Williams LD, Glenn AE, Zimeri AM, Bacon CW, Smith MA, Riley RT. Fumonisin disruption of ceramide biosynthesis in maize roots and the effects on plant development and *Fusarium verticillioides*-induced seedling disease. J Agric Food Chem. 2007;55(8):2937–2946. http://dx.doi.org/10.1021/jf0635614

62. Arias SL, Theumer MG, Mary VS, Rubinstein HR. Fumonisins: Probable role as effectors in the complex interaction of susceptible and resistant maize hybrids and *Fusarium verticillioides*. J Agric Food Chem. 2012;60(22):5667–5675. http://dx.doi.org/10.1021/jf3016333

63. Van Asch MAJ, Rijkenberg FHJ, Coutinho TA. Phytotoxicity of fumonisin B1, moniliformin, and T-2 toxin to corn callus cultures. Phytopathology. 1992;82:1330–1332. http://dx.doi.org/10.1094/Phyto-82-1330

64. Glenn AE, Zitomer NC, Zimeri AM, Williams LD, Riley RT, Proctor RH. Transformation-mediated complementation of a FUM gene cluster deletion in *Fusarium verticillioides* restores both fumonisin production and pathogenicity on maize seedlings. Mol Plant Microbe Interact. 2008;21(1):87–97. http://dx.doi.org/10.1094/MPMI-21-1-0087

65. Desjardins AE, Plattner RD, Nelsen TC, Leslie JF. Genetic analysis of fumonisin production and virulence of *Gibberella fujikuroi* mating population A (*Fusarium moniliforme*) on maize (*Zea mays*) seedlings. Appl Environ Microbiol. 1995;61(1):79–86.

66. Shim W-B, Woloshuk CP. Regulation of fumonisin B1 biosynthesis and conidiation in *Fusarium verticillioides* by a cyclin-like (C-type) gene, *FCC1*. Appl Environ Microbiol. 2001;67(4):1607–1612. http://dx.doi.org/10.1128/AEM.67.4.1607-1612.2001

67. Keyser Z, Vismer HF, Klaasen JA, Snijman PW, Marasas WFO. The antifungal effect of fumonisin B1 on *Fusarium* and other fungal species. S Afr J Sci. 1999;95(10):455–458.

68. Vesonder RF, Ciegler A, Jensen AH, Rohwedder WK, Weisleder D. Co-identity of the refusal and emetic principle from *Fusarium*-infected corn. Appl Environ Microbiol. 1976;31(2):280–285.

69. Young LG, McGirr L, Valli VE, Lumsden JH, Lun A. Vomitoxin in corn fed to young pigs. J Anim Sci. 1983;57(3):655–664.

70. Bergsjø B, Matre T, Nafstad I. Effects of diets with graded levels of deoxynivalenol on performance in growing pigs. J Vet Med. 1992;39(1–10):752–758. http://dx.doi.org/10.1111/j.1439-0442.1992.tb00240.x

71. Pestka JJ. Deoxynivalenol: Toxicity, mechanisms and animal health risks. Anim Feed Sci Technol. 2007;137(3):283–298. http://dx.doi.org/10.1016/j.anifeedsci.2007.06.006

72. Yoshizawa T, Morooka N. Acute toxicities of the new trichothecene mycotoxins: Deoxynivalenol and its monoacetate. J Food Hyg Soc Japan. 1974;15:261–269.

73. European Union. Commission regulation (EC) no 1881/2006 of 19 December 2006, as amended, on setting maximum levels for certain contaminants in foodstuffs [document on the Internet]. c2006 [cited 2015 Mar 31]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006R1881&from=EN

74. Hamilton PB. Factors influencing activity of fungi and antifungal agents in poultry feed. In: Trichothecenes and other mycotoxins. New York: John Wiley and Sons; 1985. p. 207–218.

75. Hedman R, Pettersson H, Engström B, Elwinger K. Fossum O. Effects of feeding nivalenol-contaminated diets to male broiler chickens. Poultry Sci. 1995;74(4):620–625.

76. Marasas WFO, Kellerman TS, Gelderblom WCA, Coetzer JAW, Thiel PG, Van der Lugt JJ. Leukoencephalomalacia in a horse induced by fumonisin B1, isolated from *Fusarium moniliforme*. Onderstepoort J Vet Res. 1988;55(4):197–203.

77. Kellerman TS, Marasas WFO, Thiel PG, Gelderblom WC, Cawood M, Coetzer JA. Leukoencephalomalacia in two horses induced by oral dosing of fumonisin B1. Onderstepoort J Vet Res 1990;57(4):269–275.

78. Bucci T, Hansen DK, LaBorde JB. Leucoencephalomalacia and hemorrhage in the brain of rabbits gavaged with mycotoxin fumonisins B1. Nat Toxins. 1996;4(1):51–52.

79. Colvin BM, Harrison LR. Fumonisin-induced pulmonary edema and hydrothorax in swine. Mycopathologia. 1992;117(1–2):79–82.

80. Gelderblom WCA, Jaskiewicz K, Marasas WFO, Thiel PG, Horak RM, Vleggaar R, et al. Fumonisins: Novel mycotoxins with cancer promoting activity produced by *Fusarium moniliforme*. Appl Environ Microbiol. 1988;54(7):1806–1811.

81. Gelderblom WCA, Jaskiewicz K, Marasas WFO, Thiel PG. Toxicity and carcinogenicity of the *Fusarium moniliforme* metabolite, fumonisin B1 in rats. Carcinogenesis. 1991;12(7):1247–1251. http://dx.doi.org/10.1093/carcin/12.7.1247

82. Rheeder JP, Marasas WFO, Thiel PG, Sydenham EW, Shephard GS, Van Schalkwyk DJ. *Fusarium moniliforme* and fumonisins in corn in relation to human esophageal cancer in Transkei. Phytopathology. 1992;82:353–357. http://dx.doi.org/10.1094/Phyto-82-353

83. Marasas WFO. Occurrence of *Fusarium moniliforme* and fumonisins in maize in relation to human health. S Afr Med J. 1993;83(6):382–383.

84. World Health Organization (WHO). Safety evaluation of certain mycotoxins in food. Geneva: WHO; 2001.

85. Williams JH, Grubb JA, Davis JW, Wang JS, Jolly PE, Ankrah NA, et al. HIV and hepatocellular and esophageal carcinomas related to consumption of mycotoxin-prone foods in sub-Saharan Africa. Am J Clin Nutr. 2010;92(1):154–160. http://dx.doi.org/10.3945/ajcn.2009.28761

86. Kuiper-Goodman T, Scott PM, Watanabe H. Risk assessment of the mycotoxin zearalenone. Regul Toxicol Pharmacol. 1987;7(3):253–306.

87. Stob J, Baldwin RS, Tuite J, Andrews FN, Gillette KG. Isolation of an anabolic uterotrophic compound form corn infected with *Gibberella zeae*. Nature. 1962;196:1318–1318.

88. Allen NK, Mirocha CJ, Aakus-Allen S, Bitgood JJ, Weaver G, Bates F. Effect of dietary zearalenone on reproduction of chickens. Poultry Sci. 1981;60(6):1165–1174.

89. Allen NK, Peguri A, Mirocha CJ, Newman JA. Effects of *Fusarium* cultures, T-2 toxin and zearalenone on reproduction of turkey females. Poultry Sci. 1983;62(2):282–289.

90. Zinedine A, Soriano JM, Moltó JC, Mañes J. Review on the toxicity, occurrence, metabolism, detoxification, regulations and intake of zearalenone: An oestrogenic mycotoxin. Food Chem Toxicol. 2007;45(1):1–18. http://dx.doi.org/10.1016/j.fct.2006.07.030

91. Massart F, Meucci V, Saggese G, Soldani G. High growth rate of girls with precocious puberty exposed to estrogenic mycotoxins. J Pediatr. 2008;152(5):690–695. http://dx.doi.org/10.1016/j.jpeds.2007.10.020

92. Ahamed S, Foster JS, Bukovsky A, Wimalasena J. Signal transduction through the ras/Erk pathway is essential for the mycoestrogen zearalenone-induced cell-cycle progression in MCF-7 cells. Mol Carcinog. 2001;30(2):88–98.

93. Food and Drug Administration (FDA). Guidance for industry and FDA: Advisory levels for deoxynivalenol (DON) in finished wheat products for human consumption and grains and grain by-products used for animal feed [document on the Internet]. c2010 [cited 2015 Mar 31]. Available from: http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/ucm120184.htm

94. Food and Agriculture Organization of the United Nations (FAO). Worldwide regulations for mycotoxins in food and feed in 2003 [document on the Internet]. c2004 [cited 2015 Mar 31]. Available from http://www.fao.org/docrep/007/y5499e/y5499e00.htm

95. Magan N, Olsen M. Mycotoxins in food: Detection and control. Abington: Woodhead Publishing Limited; 2004.

96. CAC/RCP 1-1969 Revision 4: Recommended international code of practice general principles of food hygiene [document on the Internet]. c2003 [cited 2015 Mar 28]. Available from: http://www.mhlw.go.jp/english/topics/importedfoods/guideline/dl/04.pdf

97. Blandino M, Pilati A, Reyneri A, Scudellari D. Effect of maize crop residue density on Fusarium head blight and on deoxynivalenol contamination of common wheat grains. Cereal Res Commun. 2010;38(4):550–559. http://dx.doi.org/10.1556/CRC.38.2010.4.12

98. Gilbert J, Tekauz A. Strategies for management of Fusarium head blight (FHB) in cereals. Prairie Soils Crops J. 2011;4:97–104.

99. McMullen M, Lukach J, McKay K, Schatz B. Fungicide and biological agent effects on Fusarium head blight across two wheat classes. J Appl Genet. 2002;43A:223–226.

100. Munkvold GP. Cultural and genetic approaches to managing mycotoxins in maize. Annu Rev Phytopathol. 2003;41:99–116. http://dx.doi.org/10.1146/annurev.phyto.41.052002.095510

101. Parsons MW, Munkvold GP. Associations of planting date, drought stress, and insects with *Fusarium* ear rot and fumonisin B1 contamination in California maize. Food Addit Contam A. 2010;27(5):591–607. http://dx.doi.org/10.1080/19440040903456337

102. Dutton MF, Westlake K. Occurrence of mycotoxins in cereals and animal feedstuffs in Natal, South Africa. J Assoc Off Anal Chem. 1985;68(5):839–842.

103. Ferreira MI, Reinhardt CG, Lamprecht SC, Sinclair M, MacKenzie L, Van Coller GJ. Morphological identification of the ryegrass hybrid *Lolium multiflorum* × *Lolium perenne* and isolation of the pathogen *Fusarium pseudograminearum* in the Western Cape. S Afr J Plant Soil. 2015;32(1):9–15.

104. Van Zyl K. The control of fungal, viral and bacterial diseases in plants. Johannesburg: Association of Veterinary and Crop Associations of South Africa; 2011.

105. Edwards SG, Pirgozliev SR, Hare MC, Jenkinson P. Quantification of trichothecene-producing *Fusarium* species in harvested grain by competitive PCR to determine the efficacy of fungicides against Fusarium head blight of winter wheat. Appl Environ Microbiol. 2001;67(4):1575–1580. http://dx.doi.org/10.1128/AEM.67.4.1575-1580.2001

106. Janse van Rensburg B, McLaren NW, Flett BC, Schoeman A. The effects of cultivar and prophylactic fungicide spray for leaf diseases on colonisation of maize ears by fumonisin producing *Fusarium* spp. and fumonisin synthesis in South Africa. Crop Prot. 2015;79:56–63. http://dx.doi.org/10.1016/j.cropro.2015.10.009

107. Small IM, Flett BC, Marasas WFO, McLeod A, Viljoen A. Use of resistance elicitors to reduce *Fusarium* ear rot and fumonisin accumulation in maize. Crop Prot. 2012;41:10–16. http://dx.doi.org/10.1016/j.cropro.2012.05.016

108. Pal KK, McSpadden-Gardener B. Biological control of plant pathogens. The Plant Health Instructor. 2006; 25 pages. http://dx.doi.org/10.1094/PHI-A-2006-1117-02

109. Schisler DA, Khan NI, Boehm MJ. Biological control of Fusarium head blight of wheat and deoxynivalenol levels in grain via use of microbial antagonists. Adv Exp Med Biol. 2002;504:53–69.

110. Diamond H, Cooke B. Preliminary studies on biological control of *Fusarium* ear blight complex of wheat. Crop Prot. 2003;22:99–107. http://dx.doi.org/10.1016/S0261-2194(02)00117-5

111. Idris HA, Labuschagne N, Korsten L. Screening rhizobacteria for biological control of *Fusarium* root and crown rot of sorghum in Ethiopia [document on the Internet]. c2007 [cited 2015 Mar 30]. Available from: http://repository.up.ac.za/bitstream/handle/2263/3201/Idris_Screening(2007).pdf?sequence=1

112. McCartney CA, Somers DJ, Fedak G, DePauw RM, Thomas J, Fox SL, et al. The evaluation of FHB resistance QTLs introgressed into elite Canadian spring wheat germplasm. Mol Breeding. 2007;20(3):209–221. http://dx.doi.org/10.1007/s11032-007-9084-z

113. Yu JB, Bai GH, Cai SB, Ban T. Marker-assisted characterization of Asian wheat lines for resistance to Fusarium head blight. Theor Appl Genet. 2006;113(2):308–320. http://dx.doi.org/10.1007/s00122-006-0297-z

114. Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica. 2005;142(1):169–196. http://dx.doi.org/10.1007/s10681-005-1681-5

115. Van Coller GJ, Lamprecht SC, Viljoen A. Resistance of South African wheat cultivars and test lines to Fusarium head blight caused by *F. graminearum* s.s. and *F. pseudograminearum*. In: 49th Congress of the Southern African Society for Plant Pathology; 2015 January 18–23; Bloemfontein, South Africa. Available from: http://www.saspp.co.za/forms/SASPP-2015-abstracts.pdf

116. Small IM, Flett BC, Marasas WFO, McLeod A, Stander MA, Viljoen A. Resistance in maize inbred lines to *Fusarium verticillioides* and fumonisin accumulation in South Africa. Plant Dis. 2012;96(6):881–888. http://dx.doi.org/10.1094/PDIS-08-11-0695

117. Robertson-Hoyt LA, Jines MP, Balint-Kurti PJ, Kleinschmidt CE, White DG, Payne GA, et al. QTL mapping for *Fusarium* ear rot and fumonisin contamination resistance in two maize populations. Crop Sci. 2006;46(4):1734–1743. http://dx.doi.org/10.2135/cropsci2005.12-0450

118. Ding J-Q, Wang X, Chander S, Yan J, Li J. QTL mapping of resistance to *Fusarium* ear rot using a RIL population in maize. Mol Breed. 2008;22(3):395–403. http://dx.doi.org/10.1007/s11032-008-9184-4

119. Rami JF, Dufour P, Trouche G, Fliedel G, Mestres C, Davrieux F, et al. Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* (L.) Moench). Theor Appl Genet. 1998;97(4):605–616. http://dx.doi.org/10.1007/s001220050936

120. Rooney WL, Klein RR. Potential of marker assisted selection for improving grain mold resistance in sorghum. In: Chandrashekar A, Bandyopadhyay R, Hall AL, editors. Technical and institutional options for sorghum grain mold management: Proceedings of an international consultation; 2000 May 18–19; Patancheru, India. Patancheru: ICRISAT; 2000. p. 183–194.

121. Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, et al. DArT markers: Diversity analyses and mapping in *Sorghum bicolor*. BMC Genomics. 2008;9:26. http://dx.doi.org/10.1186/1471-2164-9-26

122. Munkvold GP, Hellmich RL, Rice LG. Comparison of fumonisin concentrations in kernels of transgenic Bt maize hybrids and nontransgenic hybrids. Plant Dis. 1999;83(2):130–138. http://dx.doi.org/10.1094/PDIS.1999.83.2.130

123. Duvick J, Rood T, Maddox JR, Wang X. Fumonisin detoxification compositions and methods. United States Patent US5792931 [online]. c1998 [cited 2015 Mar 28]. Available from: http://www.google.com/patents/US5792931#forward-citations

124. Munkvold GP, Desjardins AE. Fumonisins in maize: Can we reduce their occurrence? Plant Dis. 1997;81(6):556–565. http://dx.doi.org/10.1094/PDIS.1997.81.6.556

125. Ono EYS, Sasaki EY, Hashimoto EH, Hara LN, Correa B, Itano EN, et al. Post-harvest storage of corn: Effect of beginning moisture content on mycoflora and fumonisin contamination. Food Addit Contam. 2002;19(11):1081–1090. http://dx.doi.org/10.1080/02652030210146828

126. Richard JL. Some major mycotoxins and their mycotoxicoses – An overview. Int J Food Microbiol. 2007;119(1–2):3–10. http://dx.doi.org/10.1016/j.ijfoodmicro.2007.07.019

127. Sydenham EW, Van der Westhuizen L, Stockenström S, Shephard GS, Thiel PG. Fumonisin-contaminated maize: Physical treatment for the partial decontamination of bulk shipments. Food Addit Contam. 1994;11(1):25–32.

128. Mogensen JM, Sørensen SM, Sulyok M, Van der Westhuizen L, Shephard GS, Frisvad JC, et al. Single kernel analysis of fumonisins and other fungal metabolites in maize from South African subsistence farmers. Food Addit Contam. 2011;28(12):1724–1734. http://dx.doi.org/10.1080/19440049.2011.611823

129. Shetty PH, Bhat RV. A physical method for segregation of fumonisin-contaminated maize. Food Chem. 1999;66(3):371–374. http://dx.doi.org/10.1016/S0308-8146(99)00052-7

130. Abbas HK, Mirocha CJ, Pawlosky RJ, Pusch DJ. Effect of cleaning, milling, and baking on deoxynivalenol in wheat. Appl Environ Microbiol. 1985;50(2):482–486.

131. Shephard GS, Leggott NL, Stockenström S, Somdyala NIM, Marasas WFO. Preparation of South African maize porridge: Effect on fumonisin mycotoxin levels. S Afr J Sci. 2002;98(7/8):393–396.

132. Trenholm HL, Charmley LL, Prelusky DB, Warner RM. Washing procedures using water or sodium carbonate solutions for the decontamination of three cereals contaminated with deoxynivalenol and zearalenone. J Agric Food Chem. 1992;40(11):2147–2151. http://dx.doi.org/10.1021/jf00023a021

133. Young JC, Subryan LM, Potts D, McLaren ME, Gobran FH. Reduction in levels of deoxynivalenol in contaminated wheat by chemical and physical treatment. J Agric Food Chem. 1986;34(3):461–465.

134. Young JC, Zhu H, Zhou T. Degradation of trichothecene mycotoxins by aqueous ozone. Food Chem Toxicol. 2006;44(3):417–424. http://dx.doi.org/10.1016/j.fct.2005.08.015

135. Abramson D, House JD, Nyachoti CM. Reduction of deoxynivalenol in barley by treatment with aqueous sodium carbonate and heat. Mycopathologia. 2005;160(4):297–301. http://dx.doi.org/10.1007/s11046-005-0087-1

136. Wilson SC, Brasel TL, Martin JM, Wu C, Andriychuk L, Douglas DR, et al. Efficacy of chlorine dioxide as a gas and in solution in the inactivation of two trichothecene mycotoxins. Int J Toxicol. 2005;24(3):181–186. http://dx.doi.org/10.1080/10915810590953437

137. Karlovsky P. Biological detoxification of the mycotoxin deoxynivalenol and its use in genetically engineered crops and feed additives. Appl Microbiol Biotechnol. 2011;91(3):491–504. http://dx.doi.org/10.1007/s00253-011-3401-5

138. Karlovsky P. Biological detoxification of fungal toxins and its use in plant breeding, feed and food production. Nat Toxins. 1999;7(1):1–23. http://dx.doi.org/10.1002/(SICI)1522-7189(199902)7:1<1::AID-NT37>3.0.CO;2-9

139. Rheeder JP, Shephard GS, Vismer HF, Gelderblom WCA. Medical Research Council (MRC) policy brief. Guidelines on mycotoxin control in South African foodstuffs: From the application of the Hazard Analysis and Critical Control Point (HACCP) system to new national mycotoxin regulations [document on the Internet]. c2009 [cited 2015 Aug 18]. Available from: http://www.mrc.ac.za/policybriefs/Mycotoxinguidelines.pdf

140. Rabie CJ, Marasas WFO, Thiel PG, Lübben A, Vleggaar R. Moniliformin production and toxicity of different *Fusarium* species from southern Africa. Appl Environ Microbiol. 1982;43(3):517–521.

141. Rabie CJ, Sydenham EW, Thiel PG, Lübben A, Marasas WFO. T-2 toxin production by *Fusarium acuminatum* isolated from oats and barley. Appl Environ Microbiol. 1986;52(3):594–596.

142. Logrieco A, Moretti A, Castella G, Kostecki M, Golinski P, Ritieni A, et al. Beauvericin production by *Fusarium* species. Appl Environ Microbiol. 1998;64(8):3084–3088.

143. Marín P, Moretti A, Ritieni A, Jurado M, Vázquez CM, González-Jaén T. Phylogenetic analyses and toxigenic profiles of *Fusarium equiseti* and *Fusarium acuminatum* isolated from cereals from southern Europe. Food Microbiol. 2012;31(2):229–237. http://dx.doi.org/10.1016/j.fm.2012.03.014

144. Hossain MT. Studies of the impact of mycoflora associated with *Oryza sativa* (rice) in South Africa [PhD thesis]. Pretoria: University of Pretoria; 2013.

145. Chelkowski J, Lew H. *Fusarium* species of Liseola Section— occurrence in cereals and ability to produce fumonisins. Microbiol Alim Nutr. 1992;10:49–53.

146. Doidge EM, Bottomley AM, Van der Planck JE, Pauer GD. A revised list of plant diseases occurring in South Africa (Science Bulletin). Pretoria: Department of Agriculture; 1953.

147. Abbas HK, Mirocha CJ, Gunther R. Mycotoxins produced by toxic *Fusarium* isolates obtained from agricultural and nonagricultural Arctic areas of Norway. Mycopathologia. 1989;105(3):143–151. http://dx.doi.org/10.1007/BF00437246

148. Kokkonen M, Ojala L, Parikka P, Jestoi M. Mycotoxin production of selected *Fusarium* species at different culture conditions. Int J Food Microbiol. 2010;143(1–2):17–25. http://dx.doi.org/10.1016/j.ijfoodmicro.2010.07.015

149. Tan DC, Flematti GR, Ghisalberti EL, Sivasithamparam K, Chakraborty S, Obanor F, et al. Mycotoxins produced by *Fusarium* species associated with annual legume pastures and 'sheep feed refusal disorders' in Western Australia. Mycotoxin Res. 2011;27(2):123–135. http://dx.doi.org/10.1007/s12550-010-0085-0

150. Di Menna ME, Lauren DR, Smith WA. An effect of incubation temperature on zearalenone production by strains of *Fusarium crookwellense*. Mycopathologia. 1991;116(2):81–86. http://dx.doi.org/10.1007/BF00436369

151. Vermeulen M, Swart WJ, Gryzenhout M. *Fusarium* species associated with decayed and healthy tissue of *Amaranthus cruentus* and three weevil species in South Africa. In: 49th Congress of the Southern African Society for Plant Pathology; 2015 January 18–23; Bloemfontein, South Africa. Available from: http://www.saspp.co.za/forms/SASPP-2015-abstracts

152. Park J, Chu FS. Immunochemical analysis of trichothecenes produced by various fusaria. Mycopathologia.1993;121(3):179–192. http://dx.doi.org/10.1007/BF01104075

153. Greenhalgh R, Levandier D, Adams W, Miller JD, Blackwell BA, McAlees AJ, et al. Production and characterization of deoxynivalenol and other secondary metabolites of *Fusarium culmorum*. J Agric Food Chem. 1986;34(1):98–102.

154. Scott PM, Abbas HK, Mirocha CJ, Lawrence GA, Weber D. Formation of moniliformin by *Fusarium sporotrichioides* and *Fusarium culmorum*. Appl Environ Microbiol. 1987;53(1):196–197.

155. Matny ON. Detection of deoxynivalenol, zearalenone and T2-toxin produced by *Fusarium* species in different cultures. Int J Agric Crop Sci. 2013;5(20):2385–2389. Available from: http://ijagcs.com/wp-content/uploads/2013/09/2385-23891.pdf

156. Schütt F, Nirenberg HI, Deml G. Moniliformin production in the genus *Fusarium*. Mycotoxin Res. 1998;14(1):35–40. http://dx.doi.org/10.1007/BF02945091

157. Rheeder JP, Marasas WFO, Nelson PE. *Fusarium globosum*, a new species from corn in southern Africa. Mycologia. 1996;88(3):509–513. http://dx.doi.org/10.2307/3760891

158. Sydenham EW, Shephard GS, Stockenström S, Rheeder JP, Marasas WFO, Van der Merwe MJ. Production of fumonisin B analogues and related compounds by *Fusarium globosum*, a newly described species from corn. J Agric Food Chem. 1997;45(10):4004–4010. http://dx.doi.org/10.1021/jf9607066

159. Shephard GS, Sewram V, Nieuwoudt TW, Marasas WFO, Ritieni A. Production of the mycotoxins fusaproliferin and beauvericin by South African isolates in the *Fusarium* section Liseola. J Agric Food Chem. 1999;47(12):5111–5115. http://dx.doi.org/10.1021/jf9903713

160. Moss MO, Thrane H. *Fusarium* taxonomy with relation to trichothecene formation. Toxicol Lett. 2004;153(1):23–28. http://dx.doi.org/10.1016/j.toxlet.2004.04.021

161. De Saeger S. Determining mycotoxins and mycotoxigenic fungi in food and feed. Amsterdam: Elsevier; 2011.

162. Nelson PE, Plattner RD, Shackelford DD, Desjardins AE. Fumonisin B1 production by *Fusarium* species other than *F. moniliforme* in Section Liseola and by some related species. Appl Environ Microbiol. 1992;58(3):984–989.

163. Magnoli CE, Saenz MA, Chiacchiera SM, Dalcero AM. Natural occurrence of *Fusarium* species and fumonisin-production by toxigenic strains from poultry feeds in Argentina. Mycopathologia. 1999;145(1):35–41. http://dx.doi.org/10.1023/A:1007053617961

164. Du Toit JJ. Root rot of young maize plants – The causal fungi. S Afr J Agric Sci. 1968;11:595–604.

165. Abbas HK, Ocamb CM, Xie W, Mirocha CJ, Shier WT. First report of fumonisin B1, B2 and B3 production by *Fusarium oxysporum* var. redolens. Plant Dis. 1995;79(6):968. http://dx.doi.org/10.1094/PD-79-0968C

166. Beev G, Denev S, Bakalova D. Zearalenone-producing activity of *Fusarium graminearum* and *Fusarium oxysporum* isolated from Bulgarian wheat. Bulg J Agric Sci. 2013;19(2):255–259.

167. Leslie JF, Plattner RD, Desjardins AE, Klittich CJR. Fumonisin B1 production by strains from different mating populations of *Gibberella fujikuroi* (Fusarium Section Liseola). Phytopathology. 1992;82(3):341–345. http://dx.doi.org/10.1094/Phyto-82-341

168. Logrieco A, Moretti A, Ritieni A, Bottalico A, Corda P. Occurrence and toxigenicity of *F. proliferatum* from preharvest maize ear rot and associated mycotoxins in Italy. Plant Dis. 1995;79(7):727–731. http://dx.doi.org/10.1094/PD-79-0727

169. Ritieni A, Moretti A, Logrieco A, Bottalico A, Randazzo G, Monti SM, et al. Occurrence of fusaproliferin, fumonisin B1, and beauvericin in maize from Italy. J Agric Food Chem. 1997;45(10):4011–4016. http://dx.doi.org/10.1021/jf9702151

170. Blaney BJ, Dodman RL. Production of zearalenone, deoxynivalenol, nivalenol, and acetylated derivatives by Australian isolates of *Fusarium graminearum* and *F. pseudograminearum* in relation to source and culturing conditions. Aust J Agric Res. 2002;53(12):1317–1326.

171. Clear RM, Patrick SK, Gaba D, Roscoe M, Turkington TK, Demeke T, et al. Trichothecene and zearalenone production, in culture, by isolates of *Fusarium pseudograminearum* from western Canada. Can J Plant Pathol. 2006;28(1):131–136. http://dx.doi.org/10.1080/07060660609507279

172. Fotso J, Leslie JF, Smith JS. Production of beauvericin, moniliformin, fusaproliferin, and fumonisins B1, B2 and B3 by ex-type strains of fifteen *Fusarium* species. Appl Environ Microbiol. 2002;68(10):5195–5197. http://dx.doi.org/10.1128/AEM.68.10.5195-5197.2002

173. Gupta S, Krasnoff SB, Underwood NL, Renwick JAA, Roberts DW. Isolation of beauvericin as an insect toxin from *Fusarium semitectum* and *Fusarium moniliforme* var. *subglutinans*. Mycopathologia. 1991;115(3):185–189. http://dx.doi.org/10.1007/BF00462223

174. Farber JM, Sanders GW, Lawrence GA, Scott PM. Production of moniliformin by Canadian isolates of *Fusarium*. Mycopathologia. 1988;101(3):187–190. http://dx.doi.org/10.1007/BF00437038

175. Chelkowski J, Zawadzki M, Zajkowski P, Logrieco A, Bottalico A. Moniliformin production by *Fusarium* species. Mycotoxin Res. 1990;6(1):41–45. http://dx.doi.org/10.1007/BF03192137

176. Bacon CW, Porter JK, Norred WP, Leslie JF. Production of fusaric acid by *Fusarium* species. Appl Environ Microbiol. 1996;62(11):4039–4043.

177. Desjardins AE, Plattner RD, Gordon TR. *Gibberella fujikuroi* mating population A and *Fusarium subglutinans* from teosinte species and maize from Mexico and Central America. Mycol Res. 2000;104(7):865–872.

178. Leslie JF, Marasas WFO, Shephard GS, Sydenham EW, Stockenström S, Thiel PG. Duckling toxicity and the production of fumonisin and moniliformin by isolates in the A and F mating populations of *Gibberella fujikuroi*. Appl Environ Microbiol. 1996;62(4):1182–1187.

179. Bryden WL, Logrieco A, Abbas HK, Porter JK, Vesonder RF, Richard JL, et al. Other significant *Fusarium* mycotoxins. In: Summerell BA, Leslie JF, Backhouse D, Bryden WL, Burgess LW, editors. *Fusarium*. St. Paul, MN: APS Press; 2001. p. 360–392.

180. Schoeman A. The diversity and population structure of *Fusarium verticillioides* in South Africa [PhD thesis]. Stellenbosch: Stellenbosch University; 2014.

181. Scauflaire J, Gourgue M, Callebaut A, Munaut F. *Fusarium temperatum*, a mycotoxin-producing pathogen of maize. Eur J Plant Pathol. 2012;133(4):911–922. http://dx.doi.org/10.1007/s10658-012-9958-8

182. Marasas WFO, Smalley EB, Hamburg JR, Strong FM. Phytotoxicity of T-2 toxin produced by *Fusarium tricinctum*. Phytopathology. 1971;61(12):1488–1491.

183. Bottalico A, Logrieco A, Ritieni A, Moretti A, Randazzo G, Corda P. Beauvericin and fumonisin B1 in preharvest *Fusarium moniliforme* maize ear rot in Sardinia. Food Addit Contam. 1995;12(4):599–607.

184. Placinta CM, D'Mello JPF, MacDonald AMC. A review of worldwide contamination of cereal grains and animal feed with *Fusarium* mycotoxins. Anim Feed Sci Technol. 1999;78(1–2):21–37. http://dx.doi.org/10.1016/S0377-8401(98)00278-8

185. Aoki T, Ward TJ, Kistler HB, O'Donnell K. Systematics, phylogeny and trichothecene mycotoxin potential of Fusarium head blight cereal pathogens. Mycotoxins. 2012;62(2):91–102. http://dx.doi.org/10.2520/myco.62.91

186. South African Gazette. Foodstuffs, Cosmetics and Disinfectants Act (54/1972): Regulations governing tolerances for fungus-produced toxins in foodstuffs: Amendment [document on the Internet]. c2016 [cited 2016 Dec 07]. Available from: https://www.greengazette.co.za/documents/national-gazette-40250-of-05-september-2016-vol-615_20160905-GGN-40250.pdf

A historical assessment of sources and uses of wheat varietal innovations in South Africa

AUTHORS:
Charity R. Nhemachena[1]
Johann Kirsten[1]

AFFILIATION:
[1]Department of Agricultural Economics, Extension and Rural Development, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Charity Nhemachena

EMAIL:
cmudombi@yahoo.com

DATES:
Received: 26 Jan. 2016
Revised: 14 June 2016
Accepted: 13 Sep. 2016

KEYWORDS:
variety improvement; wheat research; Agricultural Research Council

HOW TO CITE:
Nhemachena CR, Kirsten J. A historical assessment of sources and uses of wheat varietal innovations in South Africa. S Afr J Sci. 2017;113(3/4), Art. #2016-0008, 8 pages. http://dx.doi.org/10.17159/sajs.2017/20160008

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
Bill and Melinda Gates Foundation

We undertook a historical review of wheat varietal improvements in South Africa from 1891 to 2013, thus extending the period of previous analyses. We identified popular wheat varieties, particularly those that form the basis for varietal improvements, and attempted to understand how policy changes in the wheat sector have affected wheat varietal improvements in the country over time. The empirical analysis is based on the critical review of information from policies, the varieties bred and their breeders, the years in which those varieties were bred, and pedigree information gathered from the journal *Farming in South Africa*, sourced mainly from the National Library of South Africa and the International Maize and Wheat Improvement Center (CIMMYT) database. A database of the sources and uses of wheat varietal innovations in South Africa was developed using information from the above sources. The data, analysed using trend and graphical analysis, indicate that, from the 1800s, wheat varietal improvements in the country focused on adaptability to the production area, yield potential and stability and agronomic characteristics (e.g. tolerance to diseases, pests and aluminium toxicity). An analysis of the sources of wheat varietal improvements during the different periods indicates that wheat breeding was driven initially by individual breeders and agricultural colleges. The current main sources of wheat varietal improvements in South Africa are Sensako, the Agricultural Research Council's Small Grain Institute (ARC–SGI) and Pannar. The structural changes in the agricultural sector, particularly the establishment of the ARC–SGI and the deregulation of the wheat sector, have helped to harness the previously fragmented efforts in terms of wheat breeding. The most popular varieties identified for further analysis of cost attribution and the benefits of wheat varietal improvements were Gariep, Elands and Duzi.

Significance:

- These findings form the basis for the next analysis focusing on the attribution of the benefits and costs in terms of investment in wheat breeding in South Africa.

Introduction and background

The driving factors for investment in crop varietal innovations include the need to improve (1) yield potential, (2) resistance/tolerance to biotic and abiotic stresses and (3) nutritional and processing quality.[1,2] Greater investments in agricultural research and development (R&D), particularly varietal innovations, are necessary to increase and sustain agricultural productivity, as well as to address challenges such as poverty, food security, adaptation to climate change, increased weather variability, water scarcity and the volatility of prices in global markets.[3] The World Bank's 2008 Development Report argued that productivity gains through innovations that address increasing scarcities of land and water remain the main source of growth in agriculture and a primary source of increased food and agricultural production to feed the increasing demand. Innovations such as crop varietal improvements need to focus beyond raising productivity to addressing additional challenges such as water scarcity, risk reduction, improved product quality and environmental protection.

Du Plessis[4] reported that the first wheat production in South Africa occurred in the winter of 1652 when Jan van Riebeeck planted the first winter wheat. This development in the 1600s was the foundation of all wheat production and subsequent breeding programmes to date. Despite the first production of wheat in the 1600s, wheat varietal breeding was reported to have been established more than two centuries later in 1891.[5] The focus of wheat varietal improvements in South Africa addresses the following cultivar characteristics: adaptability to the production area, yield potential and stability, and agronomic characteristics (e.g. tolerance to diseases, pests and aluminium toxicity). The wheat varietal improvement sector consists of three main actors: the Small Grain Institute of the Agricultural Research Council (ARC–SGI; established in 1976 as the then Small Grains Centre); Sensako, established in the mid-1960s (becoming autonomous in 1999 after functioning as part of Monsanto); and Pannar (entering the wheat breeding sector in the 1990s).[5]

Periodic assessment of plant breeding is required to assess the benefits of ongoing investment to allow: (1) temporary constraints that could permanently hinder the identification of crop varietal improvements to be addressed; and (2) desirable characteristics – such as quality, quantity and environmental impact – to be identified and prioritised.[6] The main objective of this study was to undertake a historical assessment of the sources and uses of wheat varietal innovations in South African agriculture. Specifically, we focused on the historical evolution of wheat varietal improvements in the country between 1891 and 2013, including the identification of popular varieties and their history, sources and uses. This assessment complements earlier efforts by Smit et al.[5], Van Niekerk[7], De Villiers and Le Roux[8] and Stander[9], firstly by extending the period of analysis from early breeding periods in the early 1900s to 2013. Furthermore, the current empirical analysis is critical in helping to identify popular wheat varieties that have been bred and grown for long periods (particularly among current varieties in the market). These varieties form the basis for analysing wheat varietal improvements in South Africa, which is the focus of a forthcoming paper in which further analysis looks at the parental history of the selected varieties from the current analysis, and develops an empirical model for the attribution of costs and benefits of wheat varietal improvements in South Africa.

Review of commercial wheat production and breeding in South Africa

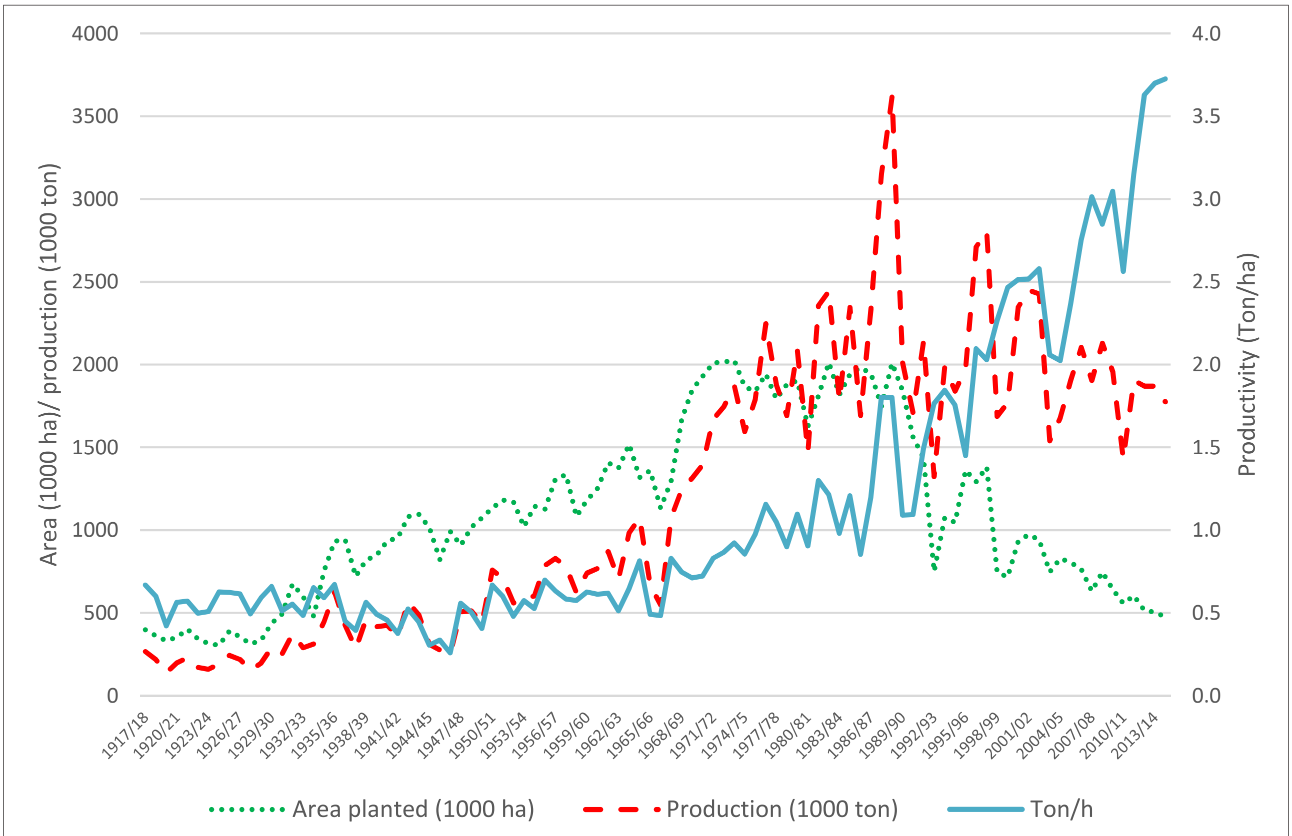
Wheat production in South Africa

The South African Department of Agriculture, Forestry and Fisheries[3] reported that the precise origin of wheat is not known, but there is evidence that the crop evolved from wild grasses somewhere in the Near East. Wheat is reported to have likely originated from the Fertile Crescent in the upper reaches of the Tigris–Euphrates drainage basin. Commercial wheat production in South Africa started in the early 1910s with varieties brought by the Dutch traders to Cape Town (then the Cape of Good Hope). Wheat is the second most important grain crop produced in South Africa after maize. In South Africa, the main uses of wheat are human consumption (especially for making flour for the bread industry), industrial (important sources of grain for alcoholic beverages, starch and straw), and animal feed (bran from flour milling as an important source of livestock feed, grain as animal feed).[3]

There are two basic types of commercially cultivated wheat in South Africa, which differ in genetic complexity, adaptation and use: (1) bread wheat (*Triticum aestivum*) and (2) durum wheat (*Triticum turgidum*). Durum wheat was derived from the fusion of two grass species some 10 000 years ago, whereas bread wheat was derived from a cross between durum wheat and a third grass species about 8000 years ago.[10] Bread wheat and durum wheat are used to make a range of widely consumed food products; for example, bread wheat is processed into leavened and unleavened breads, biscuits, cookies and noodles and durum wheat is used to make pasta (mainly in industrialised countries), as well as bread, couscous and bulgur (mainly in the developing world).[10] South Africa mainly produces bread wheat; durum wheat represents a very small percentage of the annual wheat production in the country.

Wheat is produced in 32 of South Africa's 36 crop production regions. The main wheat-producing provinces are the Western Cape (winter rainfall), Free State (summer rainfall) and Northern Cape (irrigation). Mpumalanga (irrigation) and North West (mainly irrigation) are other important wheat-producing provinces.[11] The annual wheat production in South Africa ranges from 1.5 to 3 million tonnes, with productivity rates of 2–2.5 tonnes/ha under dryland (Figure 1) and at least 5 tonnes/ha under irrigation. However, wheat production has been decreasing in recent years. Smit et al.[5] argue that efficiency, productivity and quality in wheat production has increased over time and some of the contributing factors include research efforts from various disciplines such as plant breeding, agronomy, crop physiology and crop protection. For example, the productivity levels for dryland wheat have increased from less than 0.5 tonnes/ha in 1936 to more than 3.5 tonnes/ha in 2015 (Figure 1).[12] A study by Purchase[13] reported an 87% improvement in yield and a 20% improvement in baking quality between 1930 and 1990. Local production of wheat mainly comes from the Western Cape (contributing about 650 000 tonnes), Free State (580 000 tonnes), Northern Cape (300 000 tonnes), North West (162 000 tonnes) and Mpumalanga (92 000 tonnes). South Africa is a net wheat importer and imports about 300 000 tonnes of wheat per annum.[3]

Wheat production in South Africa occurs in both summer and winter rainfall regions. Most of the production (at least 50%) happens under dryland conditions. In the summer rainfall region, at least 30% of the total harvest is produced under irrigation.[14] Production under irrigation has a higher yield potential than dryland wheat production. Dryland productivity in South Africa is very low compared to that of the major wheat-producing countries in the world. Pannar[14] attributes the 'slower than expected progress in yield increases of local breeding programmes' to stringent quality requirements for new varieties, as well as variable climatic conditions (including dry, warm winters), low soil fertility, new diseases such as yellow/stripe rust (*Puccinia striiformis*) emerging in 1996 and the emergence of new pathotypes, the introduction of the Russian wheat aphid in 1978, and a new biotype in 2005. These factors caused wheat breeding programmes to 'discontinue many promising germplasm lines'[14] despite their highly promising yield potential, as they were susceptible to new diseases and pests. The focus in wheat breeding shifted to producing varieties resistant in terms of specific



Source: South African Grain Information Services[12]

Figure 1: Commercial wheat area, production and productivity in South Africa, 1917–2015.

agronomic characteristics (e.g. pest and disease resistance) and to good-quality varieties, as opposed to high-yielding varieties.[14]

Evolution of crop production and breeding in agriculture

Various studies have reviewed the historical changes and evolution of crop production and breeding. Examples of these studies include those of Chigeza et al.[6], Byerlee and Moya[15], Heisy et al.[16], Grace and Van Staden[17] and This et al.[18] Here we briefly review these studies to understand the approaches used and some of the major findings and their implications for this paper.

In a study focusing on analysing the impact of international wheat breeding research in the developing world between 1966 and 1990, Byerlee and Moya[15] analysed the origins and trends of varieties released by national agricultural research systems (NARS) of 38 collaborating countries. The analysis of wheat varieties released by NARS included the listing of over 1300 varieties and information of their pedigrees, ecological niches and area planted. The information was used to estimate the benefits of wheat breeding on genetic yield and changes in traits such as disease resistance and quality. Byerlee and Moya[15] found an increasing proportion (84% by 1986–1990) of spring bread wheat varieties originating directly from varieties of the International Maize and Wheat Improvement Center (CIMMYT) or those with a CIMMYT parent, especially among small NARS. They also found that larger NARS used their own crosses to develop more than half of the varieties released. The analysis of wheat releases by NARS with respect to type of variety (winter bread wheat and durum wheat) and growth habit was also done for every 5-year period between 1966 and 1990. In this study we followed a similar approach to develop a comprehensive database of wheat varietal improvements in South Africa.

Smit et al.[5] summarised wheat cultivars released in South Africa between 1983 and 2008. The current study extends the analytical period to 1891 and 2013. In addition, we build on these earlier efforts to compile a comprehensive database that forms the basis for estimating the benefits and costs attributed to wheat varietal improvements in South Africa. Another addition in the current study is the provision of the institutional evolution of wheat breeding which was not included in Smit et al.'s[5] paper. Furthermore, the focus of Smit et al.[5] was more agronomic, while the current paper focuses more on the economics side of wheat-breeding developments over the study period. Also, despite listing varieties released from 1983 to 2008, Smit et al.[5] do not provide a detailed historical evolution of wheat varietal improvements in the country.

Data and research methods

The empirical analysis is based on the critical review of information from policies, varieties bred and their breeders, years when varieties were bred, and pedigree information, as gathered from the journal *Farming in South Africa*, sourced mainly from the National Library of South Africa and the CIMMYT database. The focus was to identify the sources (institutions and individuals) of wheat varietal improvement innovations; where the innovations were used (areas where the wheat varieties were grown); factors driving the innovations; and the types of wheat varietal innovations. The study analysed the wheat varieties released and/or introduced in South Africa during the period 1891–2013. A database of sources and uses of wheat varietal innovations in South Africa was developed using information from the above sources. The database shows that the wheat in South Africa has been a subject of breeding endeavours for more than two centuries, and wheat varietal improvement has rapidly expanded, particularly in the past four decades.

Based on previous studies[6,15-18], the data were analysed using trend and graphical analysis. The analysis also considered geographical region/area, as well as wheat type and growth habit. Although the database is incomplete and undoubtedly contains errors, it is to date the most comprehensive database available on the history of wheat varietal improvement in South Africa. This database will form the basis for further analysis focusing on the attribution of wheat varietal improvements in South Africa and their costs and benefits.

Liebenberg and Pardey[19] discussed historical evolution in order to document and describe major developments in the agricultural sector over the 20th and early-21st centuries and the changing policy and institutional environment of public support for agricultural R&D in South Africa. The article by Liebenberg and Pardey[19] was used to set the historical and policy context for further analysis, and the quantification and consideration of changes in public agricultural R&D investments between 1880 and 2007. We use the approach of Liebenberg and Pardey to discuss historical changes in the wheat sector and how they have shaped varietal improvements over the years.

Results and discussion

Key developments and early history of wheat varietal improvements

Wheat production was first initiated in South Africa by Jan van Riebeeck during the winter of 1652 at the Cape of Good Hope.[4] Wheat production subsequently expanded, and by 1684 there were some exports to India.[7] However, South Africa is currently a net importer of wheat. The original cultivars produced at that time originated from Europe and the East Indies, and were brought by the early settlers and trading vessels. The selection criteria for the new varieties were then focused on adaptability to the new environment, such as resistance to stem rust, periodic droughts and wind damage. Table 1 summarises the key developments (institutional and policy) throughout the history of wheat varietal improvements in South Africa from the 1600s.

The first wheat breeding programme reportedly began in 1891 in the Western Cape Province.[20] The initial series of artificial crosses between varieties was conducted in 1902 and 1904 to retain the successful resistance of Rieti wheat while replacing its poor milling quality and tendency to shed grain prior to harvesting.[7] This undertaking marked the beginning of modern wheat breeding in South Africa. The early crosses were reported to have produced only three varieties of value, which formed the basis for the first cultivars (Union, Darlvan and Nobbs) bred in South Africa. These developments preceded Neethling's first wheat breeding efforts in 1913 at the Elsenburg Research Station near Stellenbosch. The wheat breeding research efforts at the Elsenburg Research Station led to the release of at least 26 wheat varieties, which remained dominant between 1914 and 1961. Examples of the varieties released during that time include Unie17, Unie28, Unie31, Unie52 and Unie81 (all released in 1914), with Unie52 dominating between 1917 and 1927.[7]

The next release of an early wheat variety – Kleintrou (a selection from Potchefstroom Agricultural College) – occurred in 1916. Kleintrou was directly used as a parent in at least six wheat varieties (including Farrertrou, Koalisie, Stirling, Sonop and Eleksie) between 1933 and 1958. The wheat varieties known as Bobriet and Gluretty (retaining Rieti wheat as one of the parents), released in 1925, were cultivars created from crosses after this period of testing and reselection. The Gluretty variety dominated, replacing Unie52 before succumbing to rust infection and being replaced by six new varieties released in 1933 (of these, Pelgrim, Stirling and Koalisie were successful). The post-1933 breeding period included the first use of interspecific crosses, including Medeah (a *Triticum durum* wheat with excellent rust resistance, used in South Africa since about 1850) and McFadden (originating from the USA, using Marquis and Jaroslav Emmer to produce the famous Hope and H44 wheat varieties). Long-used and adapted varieties, such as Nobbs, Van Niekerk, Florence and Kleintrou, were also consistently used in the new releases. Of all the varieties released, Stirling had the greatest impact on wheat production in the Western Cape – because of its resistance to stem rust (a resistance inherited from Rieti and Medeah), as well as its adaptability and quality (inherited from Comeback) – and remained dominant until the release of Hoopvol in 1948.[7]

The Wheat Industry Control Board was established in 1935 to regulate the wheat industry in the wake of extremely poor-quality yields despite record production levels in 1935. Among the varieties available at the time, none were of bread-wheat quality. Following the establishment of the Wheat Industry Control Board, seven new varieties were released in the 1940s, with only two making an impact: Sonop (Kleintrou/Pelgrim) and Hoopvol (Kleintrou/Gluys Early/Spring Early) which became the most popular. Sonop was the first true bread wheat from the Elsenburg

Table 1: Key developments in the history of wheat varietal improvements in South Africa

Year	Historical event	Description
1652	Planting of first wheat crop by Jan van Riebeeck	First wheat planted in South Africa in the winter of 1652.
1891	First report of wheat breeding	Overman (Principal of Agricultural School, Somerset East) reported first wheat breeding programme at an agricultural institution.
1892	Breeding at original Stellenbosch Agricultural School	Blersch (Principal of original Stellenbosch Agricultural School) reported on wheat variety tests at Stellenbosch.
1898	Opening of Elsenburg School of Agriculture	Considerable amount of research conducted on wheat grain growing; unfortunately, all records of experimental work up until 1915 are missing.
1911	Establishment of Department of Agriculture	Despite undergoing various structural changes over the years, overall this Department supported and provided an institutional framework for plant breeding research (including wheat breeding).
1913	Administration of agricultural education	Three agricultural colleges established (Cedara, Potchefstroom and Grootfontein).
1913	Wheat breeding at Elsenburg Research Station	First wheat breeding programme established at Elsenburg Research Station.
1919	Establishment of Glen Agricultural College	Glen Agricultural College established.
1928	Establishment of Langgewens Cereal Station	From 1929, experiments on wheat varietal improvements were done at the Jongensklip Experimental Plot in Caledon, and from then intensive research has been conducted on wheat varieties.
1930	Establishment of Stellenbosch Glasshouse and Laboratory	The glasshouse was used for intensive rust studies, with funding from the Wheat Industry Control Board. This was extended in 1950.
1935	Establishment of Wheat Industry Control Board	The Wheat Industry Control Board was established to regulate the wheat industry (including oats, rye and barley) following poor-quality wheat production in 1935.
1958	Division of Department of Agriculture into Department of Agricultural Economics and Marketing, and Department of Agricultural Technical Services	Research function given to the Department of Agricultural Technical Services, which continued to finance research at universities.
1958	Establishment of Pannar Seeds Pty Ltd	Although established in 1958, Pannar started being involved in wheat varietal improvements in the 1990s and remains one of the three main actors in wheat breeding in South Africa.
Mid-1960s	Establishment of Sensako	Sensako became autonomous in 1999 after functioning as part of Monsanto. Sensako remains the main actor in wheat breeding in South Africa.
1975	Establishment of Small Grains Centre (SGC)	The SGC was established as a Research Centre of the Highveld Region of the Department of Agriculture. The main objective of the SGC (now the Small Grains Institute) was to help improve production of small grains, including addressing production challenges, investigating new production possibilities and transferring of information to strategic points.
1992	Establishment of Agricultural Research Council (ARC)	The establishment of the ARC in 1992 centralised all national agricultural research functions, including the mandate to serve historically segregated homeland areas.
1996	Deregulation of wheat sector	*Marketing of Agricultural Products Act, Act 47 of 1996*, which led to the deregulation of the wheat sector, has had a significant impact on both wheat research and the industry.

Sources: Various

College of Agriculture.[7] Between 1950 and 1959, four wheat varieties were released, with only two making an impact on the wheat industry: Daeraad (Unie52A/Kruger) and Dromedaris (Hope/Gluretty). Neethling's retirement and the resultant break in continuity, coupled with increased interest from his successors in terms of using wide crosses, is arguably the reason for the limited activity in terms of varietal releases during this period.

After taking over from Neethling as head of the Department of Genetics at Stellenbosch University in 1950, F.X. Laubscher introduced a completely new era in wheat breeding in the country. In 1952, he ushered in an era of extensive international cooperation with the introduction of the International Rust Nursery, in collaboration with Dr Bayless of the US Department of Agriculture. The new releases during the 1960s were based on completely new parents, combining good yield, excellent quality and disease resistance, thus surpassing the existing varieties at the time. One of the first varieties released during this period was Flameks (Mentana-Kenya-Supremo/Florence Aurore)[7], with five new varieties released between 1960 and 1970, but without the same impact on the wheat industry. New avenues in local wheat breeding were opened following the singular success of short-strawed varieties from Borlaug and CIMMYT. In the now Northern Province of South Africa, St Clair Caporn initiated wheat breeding at Potchefstroom in 1918, while Pieper started winter wheat breeding in Bethlehem in 1954 for the Free State, and Schneider started irrigation-type spring wheat breeding at Losperfontein[7] (in what is now the North West Province).

The discussion above indicates that wheat varietal improvements in the early years of wheat breeding were specific to the production area, with little or no movement from one area to another. This situation has changed over time, and wheat breeding companies – although they focus on wheat varieties specific to the different wheat-growing regions of the country – aim to produce varieties that are adaptable across the country. According to the World Bank[21], there was little movement of genetic improvement technologies in the 1950s and 1960s, especially from the temperate North to the tropical South. Their report further argues that the focus on adapting improved varieties to subtropical and tropical regions since the 1960s has generated high payoffs and pro-poor impacts, which are expected to continue to grow with rapid advances in biological and informational sciences. Byerlee and Moya[15] also found that the initial focus on CIMMYT wheat breeding activities was on specific environments (particularly irrigated areas in Mexico and South Asia), which were later expanded to rain-fed areas to incorporate resistance to diseases such as septoria (*Septoria* spp.) and stripe rust (*Puccinia striiformis* f. sp. *tritici*) into CIMMYT germplasm. The further incidence of pests and diseases challenged CIMMYT to widen the focus on resistance to pests and diseases in different environments. Similarly, in South Africa, structural changes in the agricultural sector and the liberalisation of the wheat sector have also opened up the rapid growth of wheat breeding improvements that transcend beyond original regional production areas.

Establishment of ARC–SGI and wheat varietal improvements

The ARC–SGI was established in 1975 as the Small Grains Centre. The SGI was aimed at harnessing the impact of the then-fragmented research efforts (especially small-grain breeding programmes in the then Cape, Transvaal and Orange Free State Provinces) into an organisation running along the lines of CIMMYT following the recommendation of Dr Borlaug to the Department of Agriculture.[7,8,19] The Small Grains Centre was established as a research centre of the Highveld Region of the Department of Agriculture. The main objective was to help improve the production of small grains, including addressing production challenges, investigating new production possibilities and transferring information to strategic points. The SGI became an autonomous institute on 1 April 1995.

The Wheat Board, through motivations by Dr Jos de Kock, provided funding for a new Research Building in 1989 for the centre. De Villiers and Le Roux[8] report that 90% of the infrastructure at the ARC–SGI was funded by the Wheat Board and indirectly by wheat farmers. In an effort to harness fragmented research efforts, the SGI, since its establishment in 1975, has managed to initiate the following: a national seed multiplication scheme, a national cultivar evaluation scheme and breeding of cultivars that were nationally coordinated from Bethlehem (SGI supplied at least 65% of all nationally bred cultivars up to 1996).[8] The wheat variety improvements released by the ARC–SGI were started in 1975, and its contribution to wheat breeding remains very important to South Africa. The World Bank[21] argues that in areas where markets fail and it is difficult to appropriate benefits, public investments are required in agricultural R&D, such as wheat varietal improvements.

Wheat varietal releases, sources and uses

The sources and uses of wheat varietal innovations are presented by geographical region/area and wheat growth habit. In addition, we discuss varietal improvements by wheat breeding structural/policy shifts: before the establishment of the ARC–SGI; after the establishment of ARC–SGI to deregulate the wheat sector in 1996; and post-deregulation (1997–2013).

Figure 2 shows the main wheat varietal improvement breeders for the period 1891 to 2013. Sensako has the highest number of varieties with 102 varieties since the mid-1960s. The ARC has the second highest with 51 varieties and Pannar has the third highest with 41 varieties. Monsanto follows with 20 varieties, while Professor J.H. Neethling has released 16 varieties – one of the highest number of individually released wheat varieties used to develop many South African varieties. Most of the wheat innovation in South Africa should be credited to him and his team. Analysis of the wheat varietal improvement breeders is taken further by organisation type: local private companies such as Sensako and Pannar; local public organisations such as the ARC–SGI and universities; local individuals; foreign private companies; and foreign public organisations (Figure 3). The results show that the local private sector, with a total of 171 wheat varieties, has the highest share of varieties released in South Africa for the period under study. Local public organisations, which include the ARC–SGI and universities, trail with 72 wheat varieties – less than half that of the local private sector. Results from Figures 2 and 3 clearly show that the private sector currently dominates wheat varietal improvements in the country. The current funding challenges in the public sector mean that the private sector will continue to dominate wheat varietal improvements. However, more effort would be required to support research that caters for all types of farmers, especially the emerging farmers who would want to grow wheat. This means that the public sector has a critical role to play in this area, in addition to releasing varietal innovations to large commercial farmers.

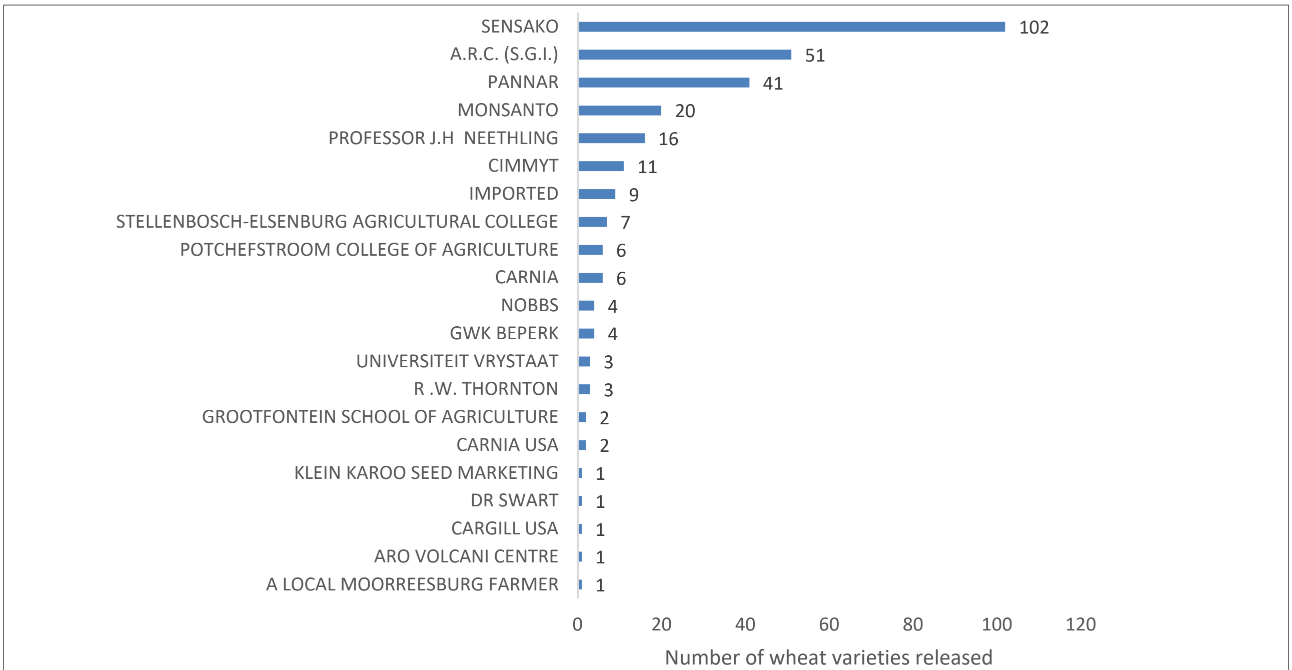


Figure 2: Number of wheat varieties released in South Africa between 1891 and 2013 by breeder.

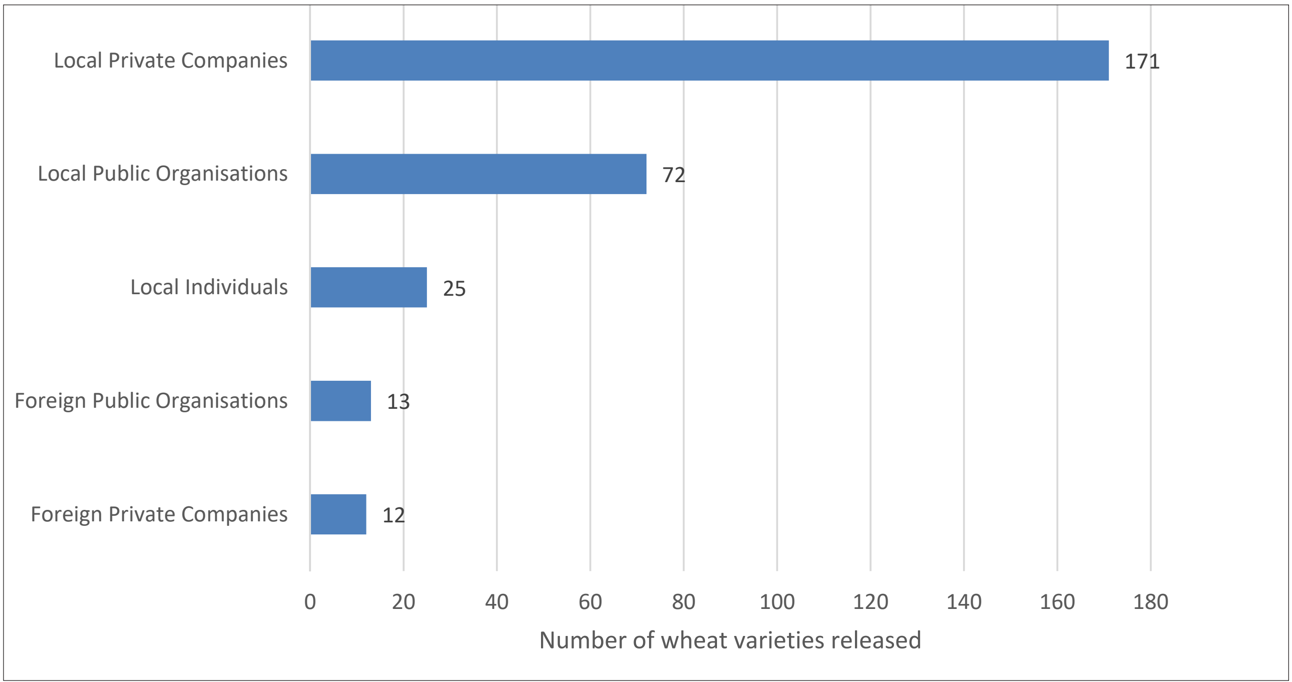


Figure 3: Classification of wheat varietal breeders in South Africa: 1891–2013.

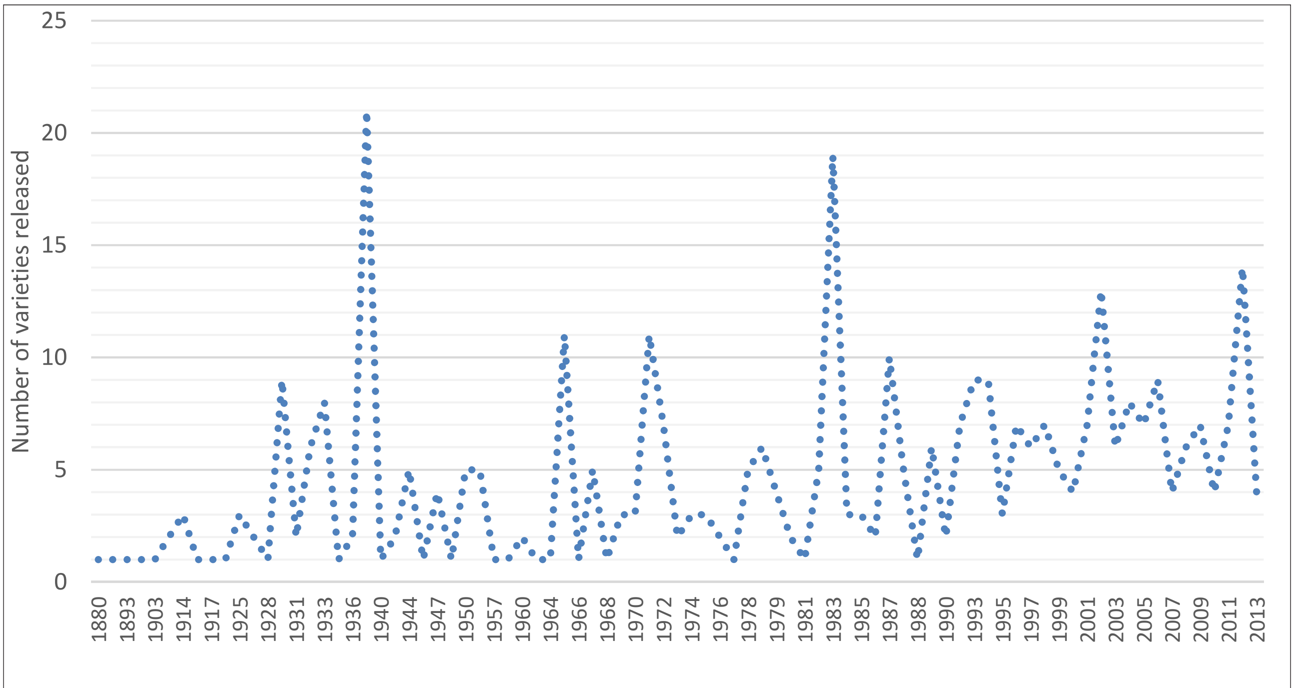


Figure 4: Number of wheat varieties released in South Africa between 1891 and 2013.

Figure 4 presents rates of varietal releases from 1880 to 2013 in South Africa. Figure 4 shows that wheat varietal innovations have evolved over time. The distinct period of interest in wheat varietal improvement history in South Africa includes pre-establishment of the Wheat Control Board in 1936; the period 1936–1976 (when the Small Grains Centre was established); 1976–1996 (when the Wheat Market was deregulated); and 1997 to the present. Prior to 1936, wheat varietal improvements and the release of varieties were driven by individual researchers and colleges of agriculture. When the Wheat Control Board was established in 1935, it started to promote research in wheat varietal innovations in the country. During these early years, funding from the Wheat Control Board was the main source of wheat varietal innovation research. The consolidation of research efforts from the then colleges of agriculture provided an important way of harnessing the synergies from the comparatively limited research capacity scattered across these colleges.[19]

The low rate of wheat varietal release in the late 1970s and early 1980s could have been driven by reduced government funding for all non-security departments in favour of increased demands for military support.[19] The introduction of the *Marketing of Agricultural Products Act* in 1996 led to the dissolution of the Wheat Board, which affected the funding originally provided by the Board for wheat varietal improvements in the country. The establishment of the ARC in 1992 centralised all national agricultural research functions, including the mandate to serve historically segregated homeland areas.

The varieties were also analysed by the geographical area for which they were released. Table 2 presents proportions of varieties released in South Africa for each geographical region. Initial wheat varietal breeding in South Africa focused on specific environments, which were expanded over time, especially since deregulation in 1996. Since the deregulation of the wheat sector, the ARC–SGI budget for wheat has steadily declined, reflecting reduced funding by the government for wheat varietal improvement.

Table 2: Proportions of varieties released in South Africa for each province

Province	Total number of varieties released (percentage)		
	1891–1975	1976–1996	1997–2013
Western Cape	50 (30.30)	9 (5.45)	10 (6.06)
Free State	26 (15.75)	9 (5.45)	6 (3.64)
Limpopo	11 (6.67)	7 (4.24)	1 (0.61)
Eastern Cape	6 (3.64)	–	–
KwaZulu-Natal	1 (0.61)	3 (1.81)	–
Mpumalanga	8 (4.85)	1 (0.61)	3 (1.82)
North West	8 (4.85)	–	–
Northern Cape	2 (1.21)	4 (2.42)	–
Total	**112 (67.88)**	**33 (20.00)**	**20 (12.12)**

Table 3 presents the distribution of wheat varieties released by growth habit. In the period 1891–1975, most of the wheat varietal releases focused on spring (13.80%); irrigation (9.85%) and winter (8.87%) growth habits. Spring (17.24%) and facultative growth habits dominated wheat varietal improvements in the period 1976 and 1996. Since the deregulation of the wheat market in 1996, spring, winter and facultative growth habits have dominated wheat varietal improvement research in South Africa. Liebenberg and Pardey[19] argue that initial agricultural R&D was decentralised and focused on specific environments and patterns of agricultural production. Specifically, the five agricultural colleges then focused their efforts on the main farming enterprises within their respective agro-ecological regions – for example, Elsenburg focused on winter grains. However, this situation has been greatly transformed over time, and public agricultural R&D is now more nationally centralised but has been experiencing a declining trend in recent years, at least since the deregulation of the wheat sector.

The analysis of wheat varietal releases was further divided into three distinct periods: the first comprising wheat varietal improvements prior to the establishment of the ARC–SGI; the second development from the establishment of the ARC–SGI to the deregulation period of the wheat sector; and the third period the post-deregulation period (1997 to 2013).

Table 3: Distribution of wheat varieties released in South Africa by growth habit

Growth habit	Total number of varieties released (percentage)		
	1891–1975	1976–1996	1997–2013
Spring	28 (13.80)	35 (17.24)	26 (12.81)
Winter	18 (8.87)	7 (3.45)	11 (5.42)
Facultative	7 (3.45)	30 (14.78)	14 (6.90)
Irrigation	20 (9.85)	3 (1.48)	4 (1.98)
Total	**73 (35.96)**	**75 (36.95)**	**55 (27.10)**

Appendix 1 of the supplementary material presents a summary of the wheat varieties released in the country, including information on their release years, breeders, last year of commercialisation, pedigrees, area suitable for planting the variety, and growth habits. The most popular ARC–SGI varieties identified for further analysis of the attribution of costs and benefits of wheat varietal improvements were Gariep, Eland and Duzi. These varieties performed well in the commercial market and were classified as some of the dominant national varieties by the South African Grain Laboratory. For example, Gariep was released in 1993 by SGI and has been one of the dominant varieties in Regions 23–25 based on the South African Grain Laboratory database of dominant varieties by region. Elands was released in 1998 by ARC–SGI and was dominant in Regions 21–26 for the period 2001–2016. This variety has remained dominant in the market for a long period. Duzi was released by ARC–SGI in 2004, and was a dominant ARC variety in Regions 30–35 for the period 2008–2015. The three varieties are of great interest because they are ARC varieties that have been dominant in the national wheat crop composition. Pedigree analysis of these varieties would be used to attribute the benefits of wheat varietal improvements among the ARC–SGI and other research agencies.

Smit et al.[5] argue that the establishment of the Wheat Board through the *Agricultural Marketing Act, Act 59 of 1968*, and the subsequent deregulation of the wheat sector in 1996 after the enactment of the *Marketing of Agricultural Products Act, Act 47 of 1996*, have had a significant impact on both wheat research and the industry. For example, during the time of the *Agricultural Marketing Act*, the Wheat Board, in addition to regulating the wheat industry, also collected levies that were used to contribute to the funding of wheat research and varietal improvements. The deregulation of the wheat sector led to a shift in most research activities to a focus on factors aimed at lowering input costs and risks while increasing the profitability of wheat production.

Conclusions and recommendations

Wheat varietal innovations are important in agriculture, as they help to improve crop productivity, adaptability and resistance to pests and diseases, and also help to protect the environment. The main objective of this paper was to examine the historical evolution of wheat varietal improvements in the country, including the identification of popular varieties, and their history, sources and uses from 1891 to 2013.

About 501 varieties were released from wheat varietal innovations in South Africa between 1891 and 2013. From the 1800s, wheat varietal improvements in the country focused on addressing: adaptability to production area; yield potential and stability; and agronomic characteristics (e.g. tolerance to diseases, pests and aluminium toxicity). The main sources of wheat varietal improvements in South Africa are Sensako, ARC–SGI and Pannar. In terms of growth habits, most wheat varietal improvements have focused on spring and winter wheat varieties grown mostly under dryland conditions. Analysis by geographical area indicates that most of the wheat varieties released between 1891 and 2013 were for the Western Cape and Free State Provinces, which are the major wheat-producing areas in the country. Wheat varietal improvements in the early years of wheat breeding were decentralised and specific to the production area, with little or no movement from area to area. The structural changes that have occurred in the agricultural sector, particularly the establishment of the ARC–SGI and the deregulation of the wheat sector, have contributed to the effort to harness the impact of the existing fragmented research efforts, especially small-grain breeding programmes in the former Cape, Transvaal and Orange Free State Provinces.

Wheat breeding was initially driven by individual breeders and agricultural colleges. Since its establishment, Sensako has been the main source of wheat varieties, followed by the ARC–SGI and Pannar. The most popular varieties identified for further analysis, in terms of the attribution of costs and benefits of wheat varietal improvements, are Gariep, Elands and Duzi. The findings from this paper form the basis for a forthcoming paper focusing on the attribution of benefits and costs in terms of investment in wheat breeding in South Africa.

Acknowledgements

This work was supported by the Bill and Melinda Gates Foundation grant to the University of Pretoria's Department of Agricultural Economics, Extension and Rural Development. The paper is part of the PhD research by C.R.N. on: 'Biological Innovation in South African Agriculture: Economics of Wheat Varietal Change, 1950–2012' which was funded from the Bill and Melinda Gates Foundation grant.

Authors' contributions

C.R.N. conceptualised the article and compiled the first draft. J.K. made conceptual contributions and refined the article.

References

1. Atack J, Coclanis P, Grantham G. Creating Abundance: Biological Innovation and American Agricultural Development – An appreciation and research agenda. Explor Econ Hist. 2009;46:160–167. http://dx.doi.org/10.1016/j.eeh.2008.11.001
2. Lantican, MA, Dubin HJ, Morris ML. Impacts of international wheat breeding research in the developing world, 1988-2002. Mexico: CIMMYT; 2005.
3. South African Department of Agriculture, Forestry and Fisheries (DAFF). Wheat production guideline. Pretoria: Directorate Plant Production, DAFF; 2010.
4. Du Plessis AJ. The history of small-grains culture in South Africa. In: Annals of the University of Stellenbosch. Vol 8. Stellenbosch: University of Stellenbosch; 1933. p. 1652–1752.
5. Smit HA, Tolmay VL, Barnard A, Jordaan JP, Koekemoer FP, Otto WM, et al. An overview of the context and scope of wheat (*Triticum aestivum*) research in South Africa from 1983 to 2008. S Afr J Plant Soil. 2010;27(1):81–96. http://dx.doi.org/10.1080/02571862.2010.10639973
6. Chigeza G, Mashingaidze K, Shanahan P. Seed yield and associated trait improvements in sunflower cultivars over four decades of breeding in South Africa. Field Crops Res. 2012;130:46–56. http://dx.doi.org/10.1016/j.fcr.2012.01.015
7. Van Niekerk H. Southern Africa wheat pool. In: Bonjean AP, Angus WJ, editors. The world wheat book: The history of wheat breeding. Paris: Lavoisier Publishing; 2001. p. 923–936.
8. De Villiers PJT, Le Roux J. Small Grain Institute: 27 years of excellence. Bethlehem, South Africa: Small Grain Institute, Agricultural Research Council; 2003.
9. Stander CJ. The economics of cultivar improvement research in the South African wheat industry [MSc thesis]. Pretoria: University of Pretoria; 2010.
10. Trethowan RM, Hodson D, Braun HJ, Pfeiffer WH, Van Ginkel M. Wheat breeding environments. In: Lantican MA, Dubin HJ, Morris ML, editors. Impacts of international wheat breeding research in the developing world, 1988-2002. Mexico: CIMMYT; 2005.
11. Southern African Grain Laboratory (SAGL). Wheat crop quality report 2011/2012 season. Pretoria: SAGL; 2012.
12. South African Grain Information Services (SAGIS). Historical information 2015 [homepage on the Internet]. No date [cited 2015 Mar 06]. Available from: http://www.sagis.org.za
13. Purchase JL, Van Lill D. Directions in breeding for winter wheat yield and quality in South Africa from 1930 to 1990. Euphytica. 1995;82(1):79–87. http://dx.doi.org/10.1007/BF00028712
14. Pannar. Wheat production guide. Greytown: Pannar; 2009.
15. Byerlee D, Moya P. Impacts of international wheat breeding research in the developing world, 1966–90. Mexico: CIMMYT; 1993.
16. Heisy PW, Lantican MA, Dubin HJ. Impacts of international wheat breeding research in the developing world, 1966–97. Mexico: CIMMYT; 2002.
17. Grace OM, Van Staden J. A horticultural history of *Lachenalia* (Hyacinthaceae). S Afr J Sci. 2003;99:526–531.
18. This P, Lacombe T, Thomas MR. Historical origins and genetic diversity of wine grapes. Trends Genet. 2006;22(9):511–519. http://dx.doi.org/10.1016/j.tig.2006.07.008
19. Liebenberg F, Pardey PG. South African agricultural R&D: Policies and public institutions, 1880-2007. Agrekon. 2011;50:11–15. http://dx.doi.org/10.1080/03031853.2011.562681
20. Neethling JH. Wheat varieties in South Africa: Their history and development until 1912. Sci Bull. 1932;108:41.
21. World Bank. World Development report 2008. Washington DC: World Bank; 2008.

The age of fossil StW573 ('Little Foot'): An alternative interpretation of $^{26}Al/^{10}Be$ burial data

AUTHORS:
Jan D. Kramers[1]
Paul H.G.M. Dirks[2]

AFFILIATIONS:
[1]Department of Geology, University of Johannesburg, Johannesburg, South Africa
[2]Department of Geosciences, College of Science and Engineering, James Cook University, Townsville, Queensland, Australia

CORRESPONDENCE TO:
Jan Kramers

EMAIL:
jkramers@uj.ac.za

DATES:
Received: 18 Mar. 2016
Revised: 16 July 2016
Accepted: 21 Sep. 2016

KEYWORDS:
burial dating; cosmogenic nuclides; muons; *Australopithecus prometheus*; burial history

HOW TO CITE:
Kramers JD, Dirks PHGM. The age of fossil StW573 ('Little Foot'): An alternative interpretation of $^{26}Al/^{10}Be$ burial data. S Afr J Sci. 2017;113(3/4), Art. #2016-0085, 8 pages. http://dx.doi.org/10.17159/sajs.2017/20160085

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa); Australian Research Council

Following the publication (Granger DE et al., Nature 2015;522:85–88) of an $^{26}Al/^{10}Be$ burial isochron age of 3.67±0.16 Ma for the sediments encasing hominin fossil StW573 ('Little Foot'), we consider data on chert samples presented in that publication to explore alternative age interpretations. ^{10}Be and ^{26}Al concentrations determined on individual chert fragments within the sediments were calculated back in time, and data from one of these fragments point to a maximum age of 2.8 Ma for the sediment package and therefore also for the fossil. An alternative hypothesis is explored, which involves re-deposition and mixing of sediment that had previously collected over time in an upper chamber, which has since been eroded. We show that it is possible for such a scenario to yield ultimately an isochron indicating an apparent age much older than the depositional age of the sediments around the fossil. A possible scenario for deposition of StW573 in Member 2 would involve the formation of an opening between the Silberberg Grotto and an upper chamber. Not only could such an opening have acted as a death trap, but it could also have disturbed the sedimentological balance in the cave, allowing unconsolidated sediment to be washed into the Silberberg Grotto. This two-staged burial model would thus allow a younger age for the fossil, consistent with the sedimentology of the deposit. This alternative age is also not in contradiction to available faunal and palaeomagnetic data.

Significance:

- Data on chert samples taken close to StW573 impose a maximum age for the fossil of 2.8 Ma – younger than the 3.67 Ma originally reported. We propose and explore a two-stage burial scenario to resolve the inconsistency and to reopen the discussion on the age of fossil StW573.

Introduction

In a recent contribution, Granger et al.[1] present ^{10}Be and ^{26}Al data on quartz from Member 2 sediments in the Silberberg Grotto in Sterkfontein Cave, South Africa, encasing StW573 ('Little Foot')[2], a complete skeleton referred to as *Australopithecus prometheus*[1]. The apparent burial isochron date of 3.67±0.16 Ma is interpreted as the age of StW573.

Almost since the discovery of StW573, its age has been a subject of controversy. Based on the concept of a laterally continuous stratigraphy for the Sterkfontein Formation[3] and a palaeomagnetic fit, an age of about 3.3 Ma was first proposed[4]. A subsequent review[5] of mainly faunal data suggested a much younger age range of between 1.5 Ma and 2.5 Ma for Member 4 at Sterkfontein as well as for the sediments encasing StW573. In a response[6], the lower age limit for Member 4 was firmly placed at ca 2 Ma (a limit since confirmed by an U-Pb age on its capping flowstone[7]) but the concerns[5] regarding an age older than 3.0 Ma for StW573 were not fully dispelled. Cosmogenic $^{26}Al/^{10}Be$ burial dating[8] then indicated that quartz in the sediments around the fossil had been underground for 4.17±0.35 Ma (later recalculated[1] to 3.94±0.20 Ma). Because of the possibility that the quartz was reworked from previous higher levels in the cave system, this date can be regarded as a maximum age.[1,7] In contrast, U-Pb dates[9] of ca 2.2 Ma on $CaCO_3$ speleothem units from below and above StW573 are minimum ages, as the dated units are not stratigraphic flowstones, but fracture fillings[10]. The new burial isochron date[1] fits within these age brackets.

The cave deposits encasing StW573 form the northwest flank of a sediment cone that occupies most of the Silberberg Grotto area and has its apex in the eastern part of the Grotto where the ceiling is highest.[11] The deposits consist mainly of matrix-supported breccia units composed of coarse-grained chert clasts and dolomite blocks, set in a muddy sand matrix that is mostly well calcified.[10] Stratification of the breccia units is illustrated by the presence of a finer-grained, weakly consolidated (possibly decalcified), clay-rich sand layer (Unit B2a in Bruxelles et al.[10]) near the base of the dated sedimentary pile, and variations in size (Unit B3[10]) and frequency (Unit B2b[10]) of chert and dolomite clasts, indicative of several fining upward cycles.[10] Fossil StW573 is embedded within Unit B2b and positioned along the stratigraphic top of this unit.[10] The layering, matrix-supported nature of the breccia deposits and clast size variations suggest that a succession of sheet-like, sand-rich debris flows deposited the composite package over a (geologically speaking) short period of time in a process similar to the deposition of the *A. sediba* skeletons at the Malapa site.[12] The age of the fossil must be similar to the depositional age of the breccia units as it is complete, articulated[13], and fully incorporated within the breccia.

In order to date StW573, Granger et al.[1] applied the burial isochron method[14-16] to the host breccia deposits. At the surface, quartz accumulates ^{10}Be and ^{26}Al by spallation reactions caused by neutrons, in turn produced in the atmosphere by cosmic rays. As the quartz is buried (whether in an alluvial or glaciogenic sediment, or a cave) it is shielded from neutrons and ^{10}Be and ^{26}Al production all but ceases (there is a much lower production rate at depth, caused by muons, discussed below). Because ^{26}Al decays faster than ^{10}Be, the $^{26}Al/^{10}Be$ abundance ratio decreases with time, allowing a burial age to be determined if their pre-burial ratio is known (this ratio can be calculated from models, or directly measured).

The data presented by Granger et al.[1] are of high quality and the interpretation appears flawless. However, the discussion on faunal data is not closed[17] and in view of the great importance of the age of StW573 in the timeline of hominin evolution, we have re-examined the data on which the 3.67±0.16 Ma burial isochron date is based, and we present an alternative interpretation that is consistent with the data, but indicative of a younger age.

Chert samples, in-situ steady-state concentrations and a maximum age

In order for a burial isochron to be useful, two conditions have to be met. First, the samples must have had, at the time of burial, a primary spread in ^{10}Be and ^{26}Al concentrations that show a correlation with each other. Second, all samples must have been buried at the same time and share the same post-burial history. The cosmogenic isotopes then decay in proportion to their abundance, so that the correlation in ^{26}Al versus ^{10}Be space persists along an isochron for which the slope decreases with time.

A primary spread of ^{10}Be and ^{26}Al concentrations with a correlation can result if (1) samples are mixtures of surface derived and previously buried quartz grains[14], or (2) material with varying residence times at the surface is sampled. In the former case, a linear correlation is expected[14], whereas the latter situation results in a gentle convex-up curve[18] as a result of ^{26}Al decaying faster than ^{10}Be. In Granger et al.[1], which is the first application of burial isochron dating to cave chronology, both factors contribute to the spread of data. Six quartz separates from mainly surface-derived bulk sediment samples (ST1, ST2, ST3, ST8, ST9 and composite STM2 dark) have relatively high ^{10}Be and ^{26}Al concentrations. Further, three chert fragments (M2CA, M2CB and M2CC) taken from the breccia immediately adjacent to the fossil and a composite (STM2 light, consisting of chert grains from bulk sediment samples ST1 and ST2) have low ^{10}Be and ^{26}Al concentrations. These are considered by Granger et al.[1] to have been derived from higher levels in the cave, at a few metres below the surface. Together, these two sample populations define the slope of the isochron and thus the age. The chert samples yielded data with exquisite precision[1], and can provide more information than just the definition of an isochron by regression. To examine this aspect it is first necessary to discuss the concentration of cosmogenic nuclides produced at depth.

Cosmogenic nuclide production rates decrease rapidly with depth under the surface, as neutron penetration in soil and rock is limited to ca 2.5 m. However, even after deeper burial, quartz still accumulates cosmogenic nuclides as a consequence of the action of muons[19,20], and this subsurface nuclide production ultimately determines the concentrations of these nuclides at depth. At constant depth underground (e.g. zero erosion), the ^{10}Be and ^{26}Al concentrations of quartz derived from the surface or higher in the cave converge to in-situ steady-state or secular equilibrium (also known as saturation) values, as the abundance of each isotope is adjusted so that its loss by radioactive decay balances its production rate P (in 10^6 atoms $g^{-1}Ma^{-1}$), which depends on depth underground. Steady state is achieved after about 10 Ma, and the relation between P values and the in-situ concentrations (atoms g^{-1}) is then:

$$[^{10}Be]_{in\ situ} = P_{10}/\lambda_{10},\ [^{26}Al]_{in\ situ} = P_{26}/\lambda_{26} \qquad \text{Equation 1}$$

where λ_{10} and λ_{26} are the respective decay constants; λ_{10}=0.4988±0.0050 Ma^{-1} and λ_{26}=0.9794±0.0230 Ma^{-1}.[1] In the more general situation in which erosion occurs, secular equilibrium is not achieved, as the shielding by overburden is steadily reduced and the nuclide production thus increases with time. For any given present-day depth, in-situ produced cosmogenic nuclide concentrations are always lower under erosion than they would be at zero erosion.

In addition to a burial age, a ^{26}Al versus ^{10}Be isochron diagram can also yield the ^{26}Al and ^{10}Be concentrations produced in situ after burial.[1] These post-burial concentrations must plot on the isochron and also on a line through the origin of the diagram with a slope corresponding to the $^{26}Al/^{10}Be$ abundance ratio of post-burial production (given by Equation 1 as ca 4.1 in steady state, although it is slightly depth dependent). This is dotted line '*s*' in Figure 1a, and its intersection with the isochron, *PI*, should represent the post-burial produced abundances: (8.5±1.3)×10^4 atoms/g of ^{26}Al and (2.1±0.3)×10^4 atoms/g of ^{10}Be (1σ uncertainties given, plotted with 2σ error bars).[1] We note that the uncertainty of *PI* is probably underestimated. The reason is that two of the samples analysed by Granger et al.[1] to define the isochron and the *PI* values are composites: 'STM2-light' consists of chert fragments from samples ST1 and ST2, and 'STM2-dark' of soil-derived, iron-oxide coated, rounded quartz grains from samples ST1, ST2, ST8 and ST9. The use of aggregates from multiple samples is common practice in measuring erosion rates, especially when the average value for a whole catchment is sought.[21] However, in a test for collinearity, this use is inappropriate as it hides any heterogeneity that may have existed.

The in-situ cosmogenic nuclide production rates at the StW573 site can also be calculated directly, using the equations of Heisinger et al.[19,20] and the empirically determined cross sections and probability factors reported by Balco et al.[15] The present depth below surface of StW573 is 23 m,[3,11] and the Silberberg Grotto above it was entirely filled by a flowstone boss and overlying breccia until this was mined out in the early 20th century. The bulk rock density used is 2.5 g/cm^3, which (with a density of 2.85 g/cm^3 for dolomite) allows for about 12% porosity if air-filled, or 8% if water-filled. An altitude correction (1500 m) was applied to the component of cosmogenic nuclides produced by stopped negative muons (accounting for between 6% and 10% of production at 23 m depth). A latitude correction, if made, would account for the deflection of cosmic rays by the earth's magnetic field at low latitudes and reduce the calculated nuclide production rates.[18,22] However, fast muons that produce the bulk of ^{10}Be and ^{26}Al at the given depth are yielded by cosmic rays with energies >20 GeV, which are unlikely to be deflected significantly. For stopped negative muons there is probably a latitude effect although it is difficult to quantify.[22] No latitude correction was applied, and the in-situ ^{10}Be and ^{26}Al concentrations obtained may be overestimated by up to a few per cent as a result.

With the above parameters, steady-state in-situ abundances of 0.0165×10^6 atoms/g ^{10}Be and 0.0678×10^6 atoms/g ^{26}Al are calculated if erosion is zero (grey diamond *calc1* in Figure 1a). If an erosion rate of 5 m/Ma (the minimum rate determined by Granger et al.[1]) is assumed, present-day in-situ abundances of 0.0120×10^6 atoms/g ^{10}Be and 0.0577×10^6 atoms/g ^{26}Al result (grey diamond *calc2* in Figure 1a). If the erosion rate was 5 m/Ma and a cave chamber had existed above the Silberberg Grotto in the past (as discussed below), calculated values would be intermediate between *calc1* and *calc2*. The higher the erosion rate, the lower the in-situ concentrations. The calculated results are, respectively, within and close to the 2σ (95% confidence) uncertainty limits of point *PI* derived from the isochron regression, which are underestimated, as discussed above. While there is thus no real contradiction between the calculated values and *PI*, the difference is nevertheless significant when reconstructing the isotope abundances of chert samples through time, as shown below.

For quartz derived from the surface or from higher levels in a cave, convergence towards the in-situ produced ^{10}Be and ^{26}Al concentrations in quartz at any depth is given by the decay law:

$$[\text{nuclide}]_t = P/\lambda + ([\text{nuclide}]_0 - P/\lambda) \times e^{-\lambda t} \qquad \text{Equation 2}$$

where t=0 denotes the present. Equation 2 can be used to calculate ^{10}Be and ^{26}Al concentrations in the past (with t being negative) and future (with t being positive) based on the measured concentrations at present. If erosion is non-zero, the process can be modelled by dividing it into time steps, each with its own depth-specific production rates.

As seen in Figure 1a, none of the chert samples have in-situ equilibrium ^{10}Be and ^{26}Al concentrations, in accord with the assessment[1] that they are derived from higher levels in the cave. Black arrows show their convergence on *PI* and grey arrows on *calc2*. Each sample has its own individual set of decay paths (calculated by using Equation 2 forward and backward in time from their measured concentrations), reflecting their individual histories within the cave. Further, although the concentrations of *PI* and *calc2* are not very different, the decay paths for the two points differ markedly from each other.

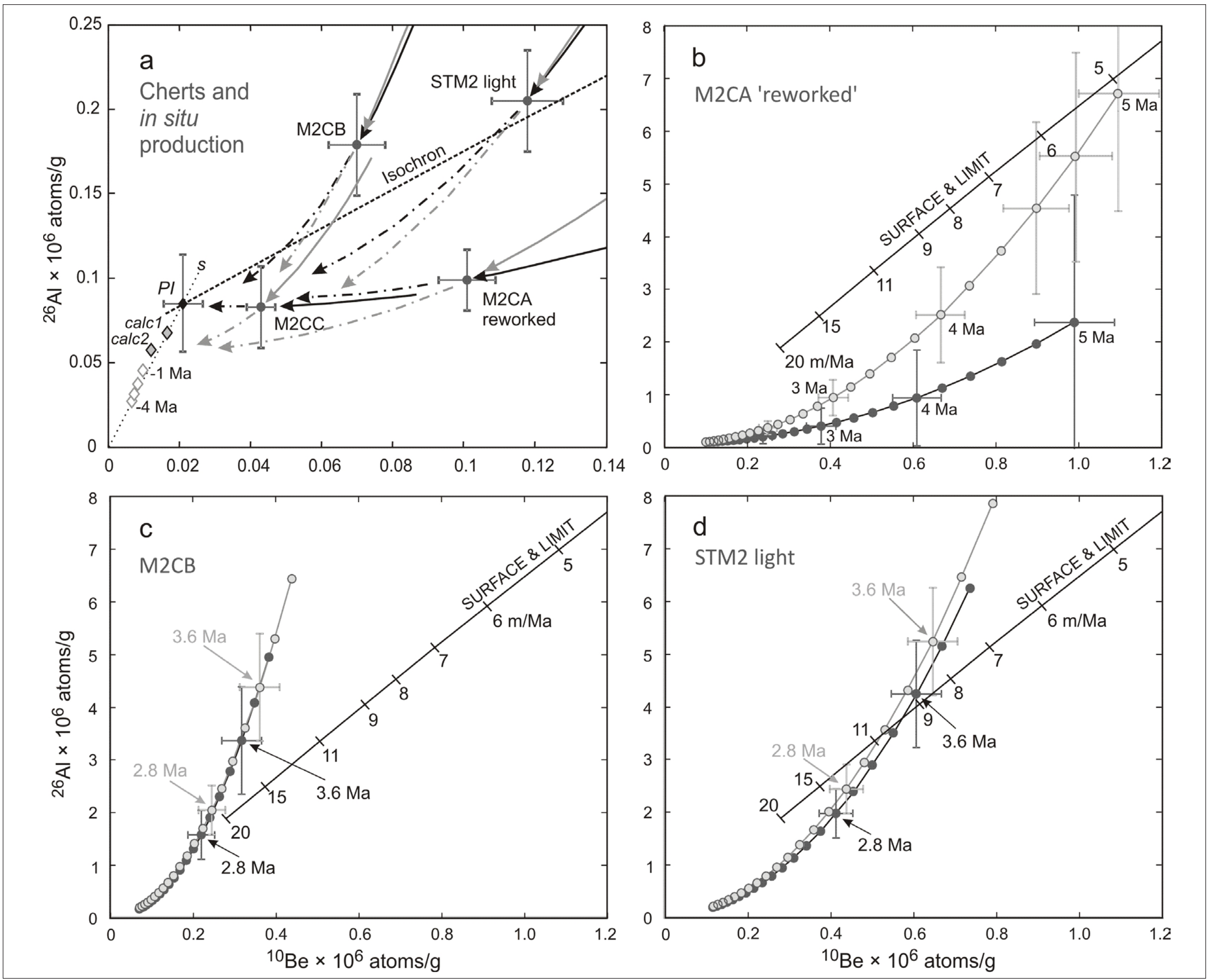


Figure 1: Chert data of Granger et al.[1] and their relationship to the in-situ produced abundances of ^{10}Be and ^{26}Al in the deposit hosting StW573. (a) Data on chert samples[1] (excluding M2CD, which plots in the 'forbidden' zone) shown with the lower part of the isochron. Post-burial in-situ produced ^{10}Be and ^{26}Al abundances must plot on or close to dotted line '*s*' (see text); its intersection with the isochron therefore yields the values of Granger et al.[1] (black diamond '*PI'*, pivot of isochron). *PI* and all chert data shown with 2σ (95% confidence) error bars. Also shown are independently calculated in-situ abundances (grey diamonds) for a depth of 23 m and average density of 2.5 g/cm^3: *calc1* shows the secular equilibrium for zero erosion and *calc2* the present day value under 5 m/Ma erosion[1], with values that quartz would have had at that location up to 4 Ma (open diamonds). The ^{10}Be and ^{26}Al concentrations of the chert samples are seen converging on in-situ points along different paths, indicating that they come from different (higher) levels in the cave system. Solid arrows depict past decay paths towards chert data as analysed, and dash–dot arrows show convergence towards the in-situ points in the future. Black arrows converge on *PI* and grey arrows on *calc2*. (b,c,d) ^{10}Be and ^{26}Al concentrations for (b) chert sample M2CA 'reworked', (c) composite chert sample MC2B and (d) chert sample STM2-light of Granger et al.[1] calculated back in time for paths corresponding to production values for *PI* (black line and symbols) and *calc* 2 (grey line and symbols). In the latter, the increase in in-situ production rates (Figure 1a) is taken into account. Error bars, shown for some ages, correspond to 2σ or 95% confidence limit. Solid line marked 'SURFACE & LIMIT' shows the steady-state abundances at the surface for erosion rates from 5 to 20 m/Ma following the surface production rates calculated by Granger et al.[1] The line also defines the upper limit for ^{26}Al/^{10}Be ratios in quartz at or below the surface.

Sample M2CA plots significantly below the isochron, and was not included in the regression of Granger et al.[1] as it was considered reworked, i.e. to come from a previous burial location in the cave system. The back-correction for sample M2CA using production rate values for the *PI* abundances (black arrows and symbols in Figure 1b), yields a curve that lies significantly (well outside 2σ, i.e. 95% confidence) below the surface production curve even at 5 Ma. This value is considered a likely maximum age for cave systems to have opened in the Cradle of Humankind UNESCO heritage site[21,23], as suggested by the absence of older fossils in the area[17,24]. A derivation – even from a few metres below the surface (which would allow a ^{26}Al/^{10}Be ratio range down to ~4.5) – is impossible for sample M2CA, because the absolute ^{10}Be abundance several million years ago would then be much lower. This mismatch suggests that the centre values for *PI* as derived from the isochron regression are inaccurate.

If M2CA is back-corrected using the parameters for *calc2* (i.e. a surface erosion rate of 5 m/Ma), the problem of its previous burial history is solved. A marginal match with near-surface abundances is achieved upward of 4.6 Ma (grey symbols and line in Figure 1b) and there is a good fit with an initial burial age of ca 5 Ma. The ^{10}Be and ^{26}Al concentrations of *calc2* correspond to the approximate upper limit for effective in-situ production rates that can provide a realistic back-correction for this sample. This result also indicates that a cave system existed at the Sterkfontein locality as early as ca 5 Ma ago, and that material reworked from this system was ultimately deposited in the Silberberg Grotto.

As the three chert samples (M2CA, M2CB and M2CC) were taken close to each other[1] (and to StW573), the same in-situ ^{10}Be and ^{26}Al production rates must have applied to all three after the sediments encasing the fossil were deposited. Using the parameters for *calc2* to examine the past of the other chert samples is, therefore, a realistic approach. Sample M2CB yields an upper age limit for the deposit that has implications for the maximum age of StW573. In Figure 1c, the back-corrected ^{10}Be and ^{26}Al concentrations for this sample are shown together with the surface production curve. ^{26}Al/^{10}Be ratios cannot plot above this curve (the 'forbidden zone'). The back-corrected values for M2CB using *calc2* production rates cross this limiting curve at 2.5 Ma, and lie within the forbidden zone outside 2σ (95% confidence) limits for ages over 2.8 Ma (grey symbols in Figure 1c). Values for 3.6 Ma clearly lie far in the forbidden zone. Sample M2CC is uninformative: it plots so close to the in-situ values that, in back-correcting, its error limits expand to include all possibilities. Values for the composite chert sample STM2-light cross the surface production curve at 3.2 Ma and move beyond 2σ uncertainty limits at 3.6 Ma (grey symbols in Figure 1d). While these values for STM2-light seem less restrictive, it must be noted that this sample is a composite and probably heterogeneous, so components of it would likely yield lower maximum ages than its bulk. Because StW573 was deposited in the Silberberg Grotto as an articulated skeleton[13], the individual either died in situ or not long before deposition. This places a maximum age constraint of ca 2.8 Ma on the fossil. The use of 95% confidence limits boosts confidence in this result.

With the recent advances in precision and accuracy of measurements of low concentrations of ^{10}Be and ^{26}Al in quartz[1] as well as a firmer basis for calculating their production rates at depth[15], the approach taken here holds promise to be useful for reconstructing the geological history of cave systems.

Exploring a two-stage burial scenario

The maximum age for the breccia deposit encasing StW573, as determined above, appears to contradict the burial isochron date of Granger et al.[1], even if the uncertainty of the latter was underestimated through the use of composite samples. This problem may be resolved by proposing that this breccia deposit contains material that was earlier buried in a chamber at a higher level in the cave system, i.e. it is a secondary deposit. An example of such a secondary deposit in Sterkfontein Cave occurs in the Name Chamber, which contains material from Member 5 (mainly) and Member 4, derived from former higher cave levels now exposed in the open excavation.[25,26] As discussed above, the breccia surrounding StW573 contains chert fragments that are derived from various levels in the cave, going back as far as about 5 Ma, indicating that these sediments were reworked. A present-day example in the Cradle of Humankind of such a two-level cave (with a potential death trap) is Gladysvale.[27]

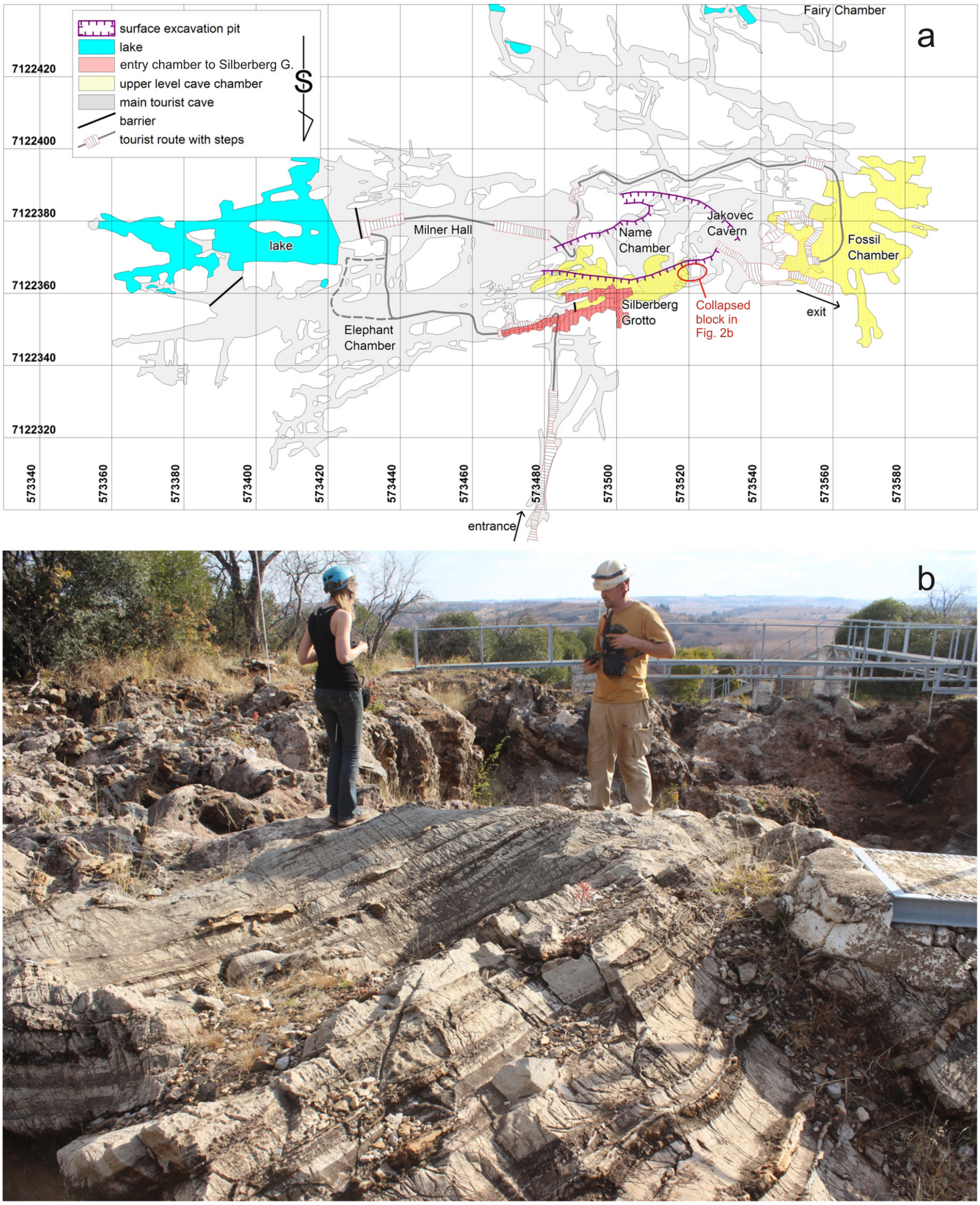
The deposits of Member 4 and 5, now exposed in the surface excavation pit, accumulated in a cave chamber between ca 2.5 and 1.4 Ma.[17] This chamber was de-roofed as a result of erosion, estimated at a rate of ca 5 m/Ma,[1] (rendering the land surface about 14 m higher at 2.8 Ma than today), and roof collapse. Figure 2a shows the position of this chamber (approximately delineated by the extent of the current excavation pit) relative to the Silberberg Grotto. Immediately south of the open excavation a large block of dolomite occurs that shows a dip of ca 30° S (Figure 2b), while the strata at Sterkfontein generally dip 25–30° NNW. This block lies above the east end of the Silberberg Grotto (Figure 2a) where the apex of its sediment cone is located.[11] It was noted by Robinson[25] as 'collapsed dolomite' but received no attention after that. This block is most likely part of a cave roof that collapsed into a void, thus documenting that a cave chamber once existed above the present Silberberg Grotto. The evidence does not allow determination of whether this chamber formed part of the large cave holding Members 4, 5 and 6, or was separate from it; but the second possibility cannot be excluded.

However, whilst a two-staged burial scenario is thus not inconsistent with the geological evidence, it must be assessed whether such a scenario could possibly result in a cosmogenic isotope array resembling an isochron. To do so, we calculated the ^{10}Be and ^{26}Al concentration data of individual samples back in time, as done for the chert samples. In Figure 3, the black symbols (here with 2σ, i.e. 95% confidence, error ellipses) and solid curves show the back-correction to 2.8 Ma for sediment samples and STM2-light, calculated using Equation 2 and applying present day in-situ abundances corresponding to *calc2* of Figure 1a. Sample ST7 of Granger et al.[1], taken at the surface and indicating an erosion rate of 5–6 m/Ma, is shown for comparison.

Although STM2-light is a composite sample, its average ^{26}Al and ^{10}Be concentrations at 2.8 Ma provide the best estimate of what in-situ accumulated cosmogenic nuclide abundances in such a previous higher level cave system could have been; at 2.8 Ma they plot just below the surface production curve (Figure 3). Long-term accumulation of ^{26}Al and ^{10}Be under shielding and with a low erosion rate (as indicated by ST7) must lead to a lower ^{26}Al/^{10}Be ratio in the sample than at the surface, as a result of the more rapid decay of ^{26}Al compared to ^{10}Be. Shielding could have many physical forms, such as overburden, or a position in a cave with a small opening. Notwithstanding the lack of constraints on actual cave configurations, cosmogenic nuclide accumulation under shielding conditions can be estimated. Various scenarios based on surface production data of Granger et al.[1], with material residing in a covered position experiencing a shielding factor that decreases from ca 99% to ca 95% over a period of ca 2 Ma, can yield ^{26}Al and ^{10}Be concentrations similar to those of 'STM2-light at 2.8 Ma' in Figure 3. This is in accord with the assessment of Granger et al.[1] that STM2-light constitutes chert debris from a higher level in the cave system. As it occurs thoroughly intermingled with material originally derived from the surface (samples ST1 and ST2), it is reasonable to conclude that the latter could also have resided at this higher level in the cave system.

Before first burial, all surface-derived samples must have had ^{10}Be and ^{26}Al concentrations plotting on the surface production curve. Given the rather large 2σ uncertainties of the back-corrected concentrations at 2.8 Ma for most samples, most of the additional correction times calculated to bring each sample back to the surface production curve also have large uncertainties. This can be illustrated by considering the varying distances from individual error ellipses to the surface production curve. For each surface-derived sample, the minimum correction time needed to intersect the surface production curve is estimated by back-correcting the point on its error ellipse closest to the surface production curve beyond 2.8 Ma, using Equation 2 (grey dot–dash curves and symbols in Figure 3). The production values corresponding to in-situ ^{10}Be and ^{26}Al steady-state concentrations of 'STM2-light at 2.8 Ma' were used for this as a best estimate. The correction times are listed for each sample in Figure 3. In a two-stage burial model, these represent the minimum residence times in the upper chamber before the samples were redeposited into their current position. It can be seen that the minimum residence times vary from 0 to 0.5 Ma (in a similar manner the maximum potential residence times can be calculated, which for all samples are >1 Ma). Note that the heterogeneity of the samples is highlighted by the surface curve intersection for composite sample STM2-dark, which reflects a higher apparent erosion rate (shorter surface residence time) than any of the bulk samples from which it was derived, indicating that the individual samples are mixtures of grain populations with different surface residence times. Interestingly, all apparent minimum pre-burial erosion rates are much lower than the erosion rate measured for today using sample ST7.[1] This difference may reflect either lower true erosion rates[21,28] or higher chemical erosion factors[29] in the past, with more of the dolomite being removed by dissolution at the surface as a consequence of a more humid climate[30].

This analysis demonstrates that an apparent isochron age of 3.67±0.16 Ma can be obtained for a secondary deposit which was laid down at a much younger age (2.8 Ma in our example), but which reworked surface-derived material that had accumulated in an upper chamber over a period as long as 1 Ma (2.8–3.8 Ma) or possibly even longer. At the same time, this observation points to a way of testing the two-staged burial hypothesis. The data array of Granger et al.[1] is technically an isochron (meaning that any scatter of the data can be the result of analytical uncertainty) because of the rather large error limits of the data on the surface-derived samples. As shown by the chert data, it should now be possible to obtain greater precision for surface-derived samples as well. If an array with greater precision on the data from surface-derived samples (and no composites) still qualifies as an isochron, then the two-staged burial hypothesis is incorrect. If there is significant scatter, it is correct.



(a) Source: Adapted from Martini et al.[41]; (b) Photo: Paul Dirks

Figure 2: Prominent surface feature at Sterkfontein and its relation to the Silberberg Grotto. (a) Cave map showing the position of surface workings, entry chambers and (b) relative to the Silberberg Grotto. (b) View from the east of a large tilted dolomite block on the south side of the open excavation, adjoining breccia of Member 4.

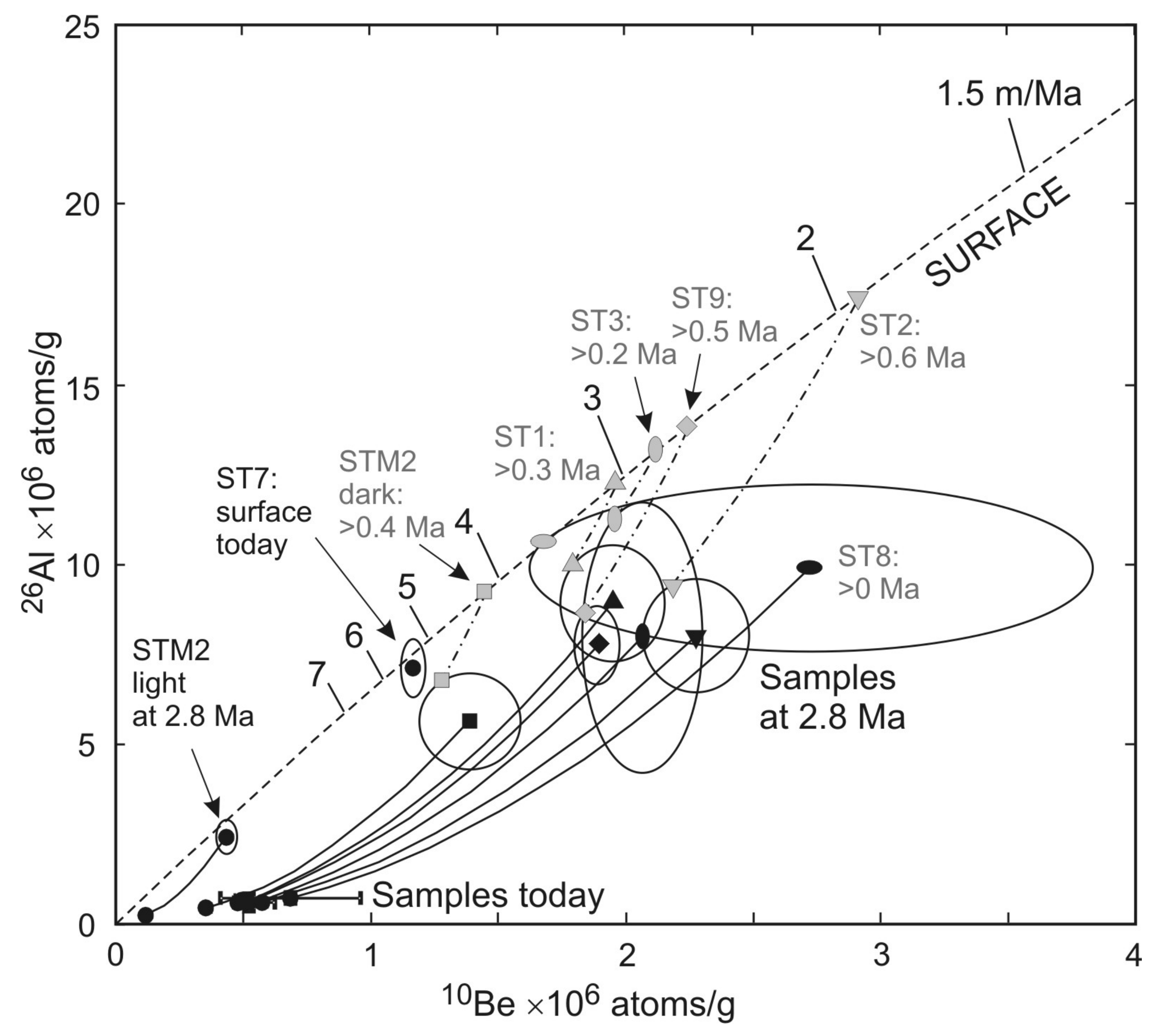


Figure 3: Back-correction of sample data in a two-stage burial model. Black symbols and solid curves show back-corrected decay paths of surface-derived samples and chert composite 'STM2-light' to 2.8 Ma, calculated using Equation 2 and in-situ ^{10}Be and ^{26}Al production rates corresponding to '*calc2*' values of Figure 1a. Error ellipses show 2σ uncertainties, derived from Granger et al.[1] Data for ST7 (present-day surface sample of Granger et al.[1]) is shown for comparison. Dashed curve 'SURFACE' shows surface steady-state ^{10}Be and ^{26}Al concentrations for a range of erosion rates. Grey dot–dash curves are decay paths in a hypothetical upper cave chamber, calculated back to their pre-burial values at surface, using in-situ ^{10}Be and ^{26}Al production rates corresponding to the concentrations of 'STM2-light at 2.8 Ma'. Minimum times required to correct back to surface values, given for each sample, yield minimum residence times in the upper chamber. Note the error ellipse for sample ST8 touches the surface curve: minimum residence time is zero for this sample.

Discussion

While we have shown that the isochron of Granger et al.[1] can be compatible with a two-stage burial scenario, the question remains as to how fossil StW573 could be younger than 2.8 million years old and be embedded in sediments that have been underground for (on average) over 3.5 Ma. In assessing possible models that fulfil the constraints imposed by the cosmogenic isotopes, our interpretation must also be consistent with the broader faunal content of sediments in the Silberberg Grotto, and palaeomagnetic results obtained from the flowstones within them (whether intrusive or stratigraphic).

To reconstruct plausible burial scenarios for StW573, it is important to assess the facies associations of the sediments surrounding the fossil, as described by Bruxelles et al.[10] These sediments are composed of surface-derived rubble, sand and mud as well as dolomite and chert fragments of varying sizes that are thoroughly mixed together.[1,10] The deposits occur as a series of layers that consist of matrix-supported breccia in which angular chert and dolomite blocks are embedded in a muddy, fine- to coarse-grained sandstone matrix with no internal structure. The clasts display a degree of grading, with variable clast sizes and clast densities across layers. The clastic sequence displays no evidence of suspension flow (e.g. cross-bedding, matrix grading, erosional channels) or standing water (e.g. mud drapes), although shelf stones show that the grotto was filled with water at times after its deposition.[11]

The deposits around StW573 have been described as the proximal to medial part of a talus cone.[11,31] The sedimentary features summarised above are consistent with the deposits being a series of sheet-like debris flows, i.e. mixtures of water, mud, sand and breccia blocks with the internal strength and ability to carry blocks (and bodies) in the matrix[32,33], yet producing preferred orientation of clasts[34]. These debris flows would have moved down the slope of a talus cone from an entry point, presumably within the roof to the eastern corner of the Silberberg Grotto.[11] The debris flow deposits display variable composition, reflecting variations in water content, provenance sediment and flow rates, but each layer was probably deposited rapidly, as demonstrated by Unit B2b[10], which envelops the fully articulated skeleton of StW573 and preserves complex body configurations of otherwise delicate elements, such as the clasped hand[13]. The rate of accumulation of the sequence as a whole cannot be determined from the sedimentology, and the isochron, being 'un-sharp', cannot constrain this aspect with any degree of confidence.

The fossil assemblage in the Silberberg Grotto preferentially comprises animals with climbing proclivities (i.e. primates and carnivores), and conspicuously lacks evidence of predator damage.[11,17] The taphonomic data indicate that many faunal remains are from individuals that entered the Silberberg Grotto on their own and were then unable to escape[35], i.e. the grotto acted as a death trap. In contrast to Member 4, which contains many hominin remains, the only hominin fossils in Member 2 are the remains of StW573, and thus the occurrence that led to a hominid entering the Silberberg Grotto appears to be rare.[35] Fossil StW573 lies embedded in Unit B2b and is thought to have been preserved in the death position,[10,13] implying that the individual died while being entombed in the debris flow, or shortly before.

When taken together, evidence suggests that StW573 ventured into an upper cave and wandered, or fell, into the Silberberg Grotto where it died and was buried. The reasons for entering the upper chamber could be many (e.g. to search for water, security, shelter), and it is plausible that the individual (like other animals in the Member 2 deposit) was unaware of the presence of the death trap, because they were unfamiliar with the cave system, or the death trap had recently formed (e.g. because part of the roof of the Silberberg Grotto had opened). Live animals falling into a death trap in such a situation can be accompanied, preceded or followed by unconsolidated sediment material that has been lying in the upper cave for hundreds of thousands of years. Erosion and re-deposition of sediment accumulations in the upper chamber would be even more likely if a passageway between the upper chamber and the Silberberg Grotto below had opened suddenly. Such a transient passageway would have disturbed the depositional environment in the upper chamber, allowing erosion, and could have created the death trap. Thus, the age for StW573 could be much younger than the cosmogenic burial age of the sediments that are now associated with the fossil in the Silberberg Grotto.

The assumption that unconsolidated sediment can be preserved in an upper chamber needs further comment. The sediment record of caves in the Cradle of Humankind site shows a significant bias towards fully lithified (i.e. calcified) sediments composed of coarser-grained, more permeable material indurated with calcite cement. In contrast, finer-grained, muddy, and less permeable material is less likely to be strongly indurated and lithified, and therefore less likely to be preserved. Yet some caves, such as the nearby Rising Star Cave[36], are known to have contained large volumes of mostly unconsolidated sediment, much of which has been eroded in response to water movement through the cave. Other examples of poorly consolidated sediment accumulations in caves include the upper flowstone-bounded units of Gladysvale with ages of up to 0.5 Ma,[27] and parts of the Member 2 deposits in the Silberberg Grotto itself (e.g. unit B2a underneath StW573[10]). Therefore, it should not come as a surprise that unconsolidated sediment may have existed for hundreds of thousands of years in an upper chamber above the Silberberg Grotto, before being washed down.

How does the burial scenario for StW573 fit with other dating constraints for sediments in the Silberberg Grotto? Palaeomagnetic work done on $CaCO_3$ units and associated siltstone material in the deposit hosting StW573 shows reverse polarity in the units below the fossil, and normal polarity at the level of the fossil and above.[37] The $CaCO_3$ units around StW573 (F2 to F4) are not flowstones, but intrusive fracture fillings.[7,10,37] According to Bruxelles et al.[10], the lower unit (F1) is also intrusive based on the presence of a void immediately above it in which botryoidal $CaCO_3$ has formed, but no real evidence is presented that F1 is not a flowstone. If the F1 unit is a stratigraphic flowstone, then the part of the deposit below StW573 would be placed in a reverse-polarity period. Given a minimum age of ca 2.2 Ma,[7,9] based on the age of the intrusive $CaCO_3$ units, and in case of a maximum age of 2.8 Ma as discussed above, the Matuyama C2r.2r Chron (2.58–2.16 Ma) would then be the only candidate[37]. If all $CaCO_3$ units are intrusive, the palaeomagnetic data have no bearing on the age of the fossil.

The fauna in the deposits of the Silberberg Grotto is largely a subset of that in Member 4 of Sterkfontein and is not highly diagnostic for age.[5,17] The fauna includes two taxa of extinct hunting hyena, *Chasmaporthetis nitidula* and *Ch. silberbergi*[38], and in the former, a similarity in primitive dentition to *Ch. australis* from the lower Pliocene fossil deposit of Langebaanweg is noted – 'although it is not clear at this stage that the two are conspecific'[38]. However, both taxa also occur in Member 4 of Sterkfontein[17], as well as in Member 1 of the Swartkrans site[24]. Member 4 has been reliably dated to between ca 2.6 Ma and 2.0 Ma by U-Pb on flowstones[7]; and at Swartkrans, $^{26}Al/^{10}Be$ burial ages from Member 1 sediments concur with U-Pb ages of flowstones between ca 2.2 and 1.8 Ma[39,40]. On the other hand, the extinct colobine monkey *Cercopithecoides williamsi*, found in the Silberberg Grotto deposits[17] as well as in Member 4 and Swartkrans Member 1[5,24], is noted as not having been reported from reliably dated sites older than 2.5 Ma[5]. In summary, no contradiction arises from these faunal data in the case of an age <2.8 Ma for the deposit encasing StW 573.

Conclusion

Cosmogenic ^{10}Be and ^{26}Al data on chert fragments from a cave deposit can impose constraints on the age of that deposit. In the case of the sediments encasing StW573, such data[1] indicate that this deposit was formed no earlier than 2.8 Ma, even if its components had been underground for (variably) longer periods, yielding an isochron age of 3.67 ± 0.16 Ma.[1] The younger age is not in conflict with faunal studies[5,17,24], palaeomagnetic work[37] and U-Pb dating[7,9]. The apparent contradiction can be resolved by invoking a two-stage burial scenario, which is geologically realistic. This scenario can ultimately yield an isochron-like data array even if primary burial ages differ among samples. It requires (1) an upper cave level environment in which sediment accumulated over time, and (2) events in which the accumulated sediment matter, including chert fragments derived from within the cave, dropped to a deeper level in the form of debris flows and was chaotically mingled. Because the fossil was incorporated as an articulated skeleton, it cannot be older than the deposit, and the individual must, therefore, have fallen into the lower cave either on its own, or incorporated in a debris flow. As the two-stage burial scenario can reconcile the indicated 2.8 Ma maximum age for the fossil with the much older isochron date, it deserves serious consideration.

Acknowledgements

We thank Darryl Granger for open discussions, fair comment and constructive criticism on earlier versions of this manuscript. Four anonymous reviewers are thanked for critical comments that led to improvements in the manuscript. J.D.K. thanks the South African National Research Foundation for incentive funding (rated researchers) and P.H.G.M.D. acknowledges funding received from the Australian Research Council (DP140104282) in support of this research.

Authors' contributions

J.D.K. contributed the considerations and calculations relating to cosmogenic nuclide systematics. P.H.G.M.D. contributed the sedimentological and taphonomical review and arguments. Both authors wrote their respective parts of the manuscript.

References

1. Granger DE, Gibbon RJ, Kuman K, Clarke RJ, Bruxelles L, Caffee MW. New cosmogenic burial ages for Sterkfontein Member 2 *Australopithecus* and Member 5 Oldowan. Nature. 2015;522:85–88. http://dx.doi.org/10.1038/nature14268
2. Clarke RJ. First ever discovery of a well-preserved skull and associated skeleton of *Australopithecus*. S Afr J Sci. 1998;94:460–463.
3. Partridge TC, Watt IB. The stratigraphy of the Sterkfontein hominid deposit and its relationship to the underground cave system. Palaeont Afr. 1991;28:35–40.
4. Partridge TC, Shaw J, Heslop D, Clarke RJ. The new hominid skeleton from Sterkfontein, South Africa: Age and preliminary assessment. J Quat Sci. 1999;14:192–198. http://dx.doi.org/10.1002/(SICI)1099-1417(199907)14:4<293::AID-JQS471>3.0.CO;2-X

5. Berger LR, Lacruz R, De Ruiter DJ. Revised age estimates of *Australopithecus*-bearing deposits at Sterkfontein, South Africa. Am J Phys Anthropol. 2002;119:192–197. http://dx.doi.org/10.1002/ajpa.10156

6. Clarke RJ. On the unrealistic 'Revised age estimates' for Sterkfontein. S Afr J Sci. 2002;98:415–418.

7. Pickering R, Kramers JD. Re-appraisal of the stratigraphy and determinations of new U-Pb dates for the Sterkfontein hominin site, South Africa. J Hum Evol. 2010;59:70–86. http://dx.doi.org/10.1016/j.jhevol.2010.03.014

8. Partridge TC, Granger DE, Caffee MW, Clarke RJ. Lower Pliocene hominid remains from Sterkfontein. Science. 2003;300:607–612. http://dx.doi.org/10.1126/science.1081651

9. Walker J, Cliff RA, Latham AG. U-Pb isotopic age of the StW 573 hominid from Sterkfontein, South Africa. Science. 2006;314:1592–1594. http://dx.doi.org/10.1126/science.1132916

10. Bruxelles L, Clarke RJ, Maire R, Ortega R, Stratford D. Stratigraphic analysis of the Sterkfontein StW 573 *Australopithecus* skeleton and implications for its age. J Hum Evol. 2014;70:36–48. http://dx.doi.org/10.1016/j.jhevol.2014.02.014

11. Clarke RJ. A deeper understanding of the stratigraphy of Sterkfontein fossil hominid site. Trans R Soc S Afr. 2006;61:111–120. http://dx.doi.org/10.1080/00359190609519960

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

13. Clarke RJ. Newly revealed information on the Sterkfontein Member 2 *Australopithecus* skeleton. S Afr J Sci. 2002;98:523–526.

14. Balco G, Rovey CWII. An isochron method for cosmogenic-nuclide dating of buried soils and sediments. Am J Sci. 2008;308:1083–1114. http://dx.doi.org/10.2475/10.2008.02

15. Balco G, Soreghan GS, Sweet DE, Marra KR, Bierman PR. Cosmogenic-nuclide burial ages for Pleistocene sedimentary fill in Unaweep Canyon, Colorado, USA. Quat Geochronol. 2013;18:149–157. http://dx.doi.org/10.1016/j.quageo.2013.02.002

16. Erlanger ED, Granger DE, Gibbon RJ. Rock uplift rates in South Africa from isochron burial dating of fluvial and marine terraces. Geology. 2012;40:1019–1022. http://dx.doi.org/10.1130/G33172.1

17. Reynolds SC, Kibii JM. Sterkfontein at 75: Review of palaeoenvironments, fauna, dating and archaeology from the hominin site of Sterkfontein (Gauteng Province, South Africa). Palaeont Afr. 2011;46:59–88.

18. Lal D. Cosmic ray labeling of erosion surfaces: *In situ* nuclide production rates and erosion models. Earth Planet Sci Lett. 1991;104:424–439. http://dx.doi.org/10.1016/0012-821X(91)90220-C

19. Heisinger B, Lal D, Jull AJT, Kubik P, Ivy-Ochs S, Neumaier S, et al. Production of selected cosmogenic radionuclides by muons 1: Fast muons. Earth Planet Sci Lett. 2002;200:345–355. http://dx.doi.org/10.1016/S0012-821X(02)00640-4

20. Heisinger B, Lal D, Jull AJT, Kubik P, Ivy-Ochs S, Knie K, et al. Production of selected radionuclides by muons 2: Capture of negative muons. Earth Planet Sci Lett. 2002;200:357–369. http://dx.doi.org/10.1016/S0012-821X(02)00641-6

21. Dirks PHGM, Placzek C, Fink D, Dosseto A, Roberts EM. Using ^{10}Be dating to reconstruct the evolving landscape in the Cradle of Humankind, South Africa. J Hum Evol. 2016;96:19–34. http://dx.doi.org/10.1016/j.jhevol.2016.03.002

22. Desilets D, Zreda M. On scaling cosmogenic nuclide production rates for altitude and latitude using cosmic-ray measurements. Earth Planet Sci Lett. 2001;193:213–225. http://dx.doi.org/10.1016/S0012-821X(01)00477-0

23. Partridge TC. Geomorphological dating of cave opening at Makapansgat, Sterkfontein, Swartkrans and Taung. Nature. 1973;246:75–79. http://dx.doi.org/10.1038/246075a0

24. McKee JK. Further chronological seriations of southern African Pliocene and Pleistocene mammalian fauna assemblages. Palaeont Afr. 1995;32:11–16.

25. Robinson JT. Australopithecines and artefacts at Sterkfontein. Part One: Sterkfontein stratigraphy and the significance of the extension site. S Afr Archaeol Bull. 1962;17:87–107. http://dx.doi.org/10.2307/3886942

26. Val A, Stratford DJ. The macrovertebrate fossil assemblage from the Name Chamber, Sterkfontein: Taxonomy, taphonomy and implications for site formation processes. Palaeont Afr. 2015;50:1–17.

27. Pickering R, Hancock PJ, Lee-Thorp JA, Grün R, Mortimer GE, McCulloch M, et al. Stratigraphy, U-Th chronology, and paleoenvironments at Gladysvale Cave: Insights into the climatic control of South African hominin-bearing cave deposits. J Hum Evol. 2007;53:602–619. http://dx.doi.org/10.1016/j.jhevol.2007.02.005

28. Dirks PHGM, Berger LR. Hominin-bearing caves and landscape dynamics in the Cradle of Humankind, South Africa. J Afr Earth Sci. 2013;78:109–131. http://dx.doi.org/10.1016/j.jafrearsci.2012.09.012

29. Riebe CS, Granger DE. Quantifying effects of deep and near-surface chemical erosion on cosmogenic nuclides in soils, saprolite, and sediment. Earth Surf Process Landforms. 2013;38:523–533. http://dx.doi.org/10.1002/esp.3339

30. Bobe R, Behrensmeyer AK. The expansion of grassland ecosystems in Africa in relation to mammalian evolution and the origin of the genus *Homo*. Palaeogeogr Palaeocl. 2004;207:399–420. http://dx.doi.org/10.1016/j.palaeo.2003.09.033

31. Stratford D, Grab S, Pickering TR. The stratigraphy and formation history of fossil- and artefact-bearing sediments in the Milner Hall, Sterkfontein Cave, South Africa: New interpretations and implications for palaeoanthropology and archaeology. J Afr Earth Sci. 2014;96:155–167. http://dx.doi.org/10.1016/j.jafrearsci.2014.04.002

32. Coussot P, Meunier M. Recognition, classification and mechanical description of debris flows. Earth Sci Rev. 1996;40:209–227. http://dx.doi.org/10.1016/0012-8252(95)00065-8

33. Iverson RM. The physics of debris flows. Rev Geophys. 1997;35:245–296. http://dx.doi.org/10.1029/97RG00426

34. Major JJ. Pebble orientation on large, experimental debris-flow deposits. Sed Geol. 1998;117:151–164. http://dx.doi.org/10.1016/S0037-0738(98)00014-1

35. Pickering TR, Clarke RJ, Heaton JL. The context of Stw573, an early hominid skull and skeleton from Sterkfontein Member 2: Taphonomy and paleoenvironment. J Hum Evol. 2004;46:279–297. http://dx.doi.org/10.1016/j.jhevol.2003.12.001

36. Dirks PHGM, Berger LR, Roberts EM, Kramers JD, Hawks J, Randolph-Quinney PS, et al. Geological and taphonomic context for the new hominin species *Homo naledi* from the Dinaledi Chamber, South Africa. eLIFE. 2015;4, Art. e09561, 37 pages. http://dx.doi.org/10.7554/eLife.09561

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

38. Turner A. Further remains of carnivore (Mammalia) from the Sterkfontein hominid site. Palaeont Afr. 1997;34:115–126.

39. Gibbon RJ, Pickering TR, Sutton MB, Heaton JL, Kuman K, Clarke RJ, et al. Cosmogenic nuclide burial dating of hominin-bearing Pleistocene cave deposits at Swartkrans, South Africa. Quat Geochron. 2014;24:10–15. http://dx.doi.org/10.1016/j.quageo.2014.07.004

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

41. Martini JEJ, Wipplinger PE, Moen HFG, Keyser A. Contribution to the speleology of Sterkfontein Cave, Gauteng Province, South Africa. Int J Speleo. 2003;32:43–69. http://dx.doi.org/10.5038/1827-806X.32.1.4

Developments in the production of economics PhDs at four research-intensive universities in South Africa

AUTHORS:
Phillip de Jager[1]
Liezel Frick[2]
Pieter van der Spuy[3]

AFFILIATIONS:
[1]Department of Finance and Tax, University of Cape Town, Cape Town, South Africa
[2]Centre for Higher and Adult Education, Stellenbosch University, Stellenbosch, South Africa
[3]Department of Accounting, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Phillip de Jager

EMAIL:
phillip.dejager@uct.ac.za

DATES:
Received: 12 Apr. 2016
Revised: 04 Sep. 2016
Accepted: 22 Sep. 2016

KEYWORDS:
doctorate; thesis; supervision; PhD by publication; economics education

HOW TO CITE:
De Jager P, Frick L, Van der Spuy P. Developments in the production of economics PhDs at four research-intensive universities in South Africa. S Afr J Sci. 2017;113(3/4), Art. #2016-0117, 9 pages. http://dx.doi.org/10.17159/sajs.2017/20160117

ARTICLE INCLUDES:
× Supplementary material
✓ Data set

FUNDING:
University of Cape Town

There is a national drive to increase PhD production, yet we know little about how this imperative takes shape within different disciplines. We therefore set out to explore recent developments and the current status of the PhD in economics at four South African research-intensive universities. A data set of all economics PhDs produced in these commerce faculties during the period 2008–2014 was analysed to determine whether the departments of economics responded to the call for increased doctoral production, and the role the PhD by publication might have played in the process. How an increase in quantity might influence doctoral education in the respective academic departments was also considered by supplementing the quantitative data with perspectives from heads of department at the four institutions. The notable increase in doctoral production over the time period studied shows that national and international trends have influenced doctoral education in economics departments within South African research-intensive universities. Increased usage of the PhD by publication has implications for policy and pedagogical practice within these departments, especially as there seems to be limited available supervisory capacity. Other changes in departmental practices, such as the entrenchment of a research culture and the promotion of collaborative research amongst students and staff, also contributed to maintain quality in doctoral education.

Significance:

- A substantial increase in the quantity of economics PhDs produced was accompanied by an unexpected increase in quality.
- The increase in quality related to management changes, including a move to the PhD by publication, increased attention to ensuring the quality of students allowed entry to PhD programmes, facilitation of full-time doctoral studies through funding arrangements, and the appointment of international faculty with a research orientation.

Introduction

> *Tradeoffs have long been at the centre of economics. The aphorism 'there is no such thing as a free lunch' captures a central economic idea: you cannot get something for nothing. Among the many tradeoffs emphasised by economists are guns v butter, public v private, efficiency v equity,* ***quality v quantity*** *or cost and short-term v long-term performance* [emphasis added].[1]

The quality versus quantity debate currently lies at the centre of national debates on the production of PhDs. The Academy of Science of South Africa (ASSAf) report[2] was the first comprehensive report on PhD training in South Africa. It covered the period from 2000 to 2007 and called for an urgent increase in the quantity of 'high-quality PhDs' produced in South Africa. The call for a national increase in doctoral production is echoed by the National Planning Commission of South Africa that envisions the nearly threefold increase in production of doctorates to 5000 doctoral graduates per year by 2030, while the National Development Plan of South Africa envisages that 70% of all academic staff employed at universities in South Africa will have doctorates within the same period of time (nearly double the current 40%).[3] The figures seem to emphasise quantity over quality, which has raised concerns over a production-focused mode of doctoral education within the academic community.[4-8]

It is not only the South African context that is important for this study. Internationally, universities are stepping away from their traditional role as leaders of knowledge production to being profit-seeking businesses. A finance-driven approach to university governance and practice underscores this trend globally.[9] Academics as cogs in a research and teaching production line militate against 'passionate scholarship' and research runs the risk of becoming repetitive and mundane.[10] PhD policies, pedagogical practices, and the eventual outputs produced – including the chosen thesis format – are arguably influenced by these trends. It is thus not surprising that there has been a notable diversification in PhD-related policies, pedagogical approaches and formats across and within institutions and disciplines. One such noteworthy change is a shift away from the traditional monograph PhD to the PhD by publication. The numbers of PhDs by publication are growing internationally.[11] Boud and Lee[12] note 'the rapidly expanding doctorates by publications that are a visible response to policy-led pressures for research productivity within the "performative" university'. A major reason for the attractiveness of this path to the PhD is higher education funding models that reward both publications and research student completion[13] – as is the case in South Africa.

However, South Africa sports an unequal institutional playing field in terms of producing doctorates and therefore it becomes difficult (and even unfair) to compare the doctoral production at all the higher education institutions. The ASSAf report[2] found that the top nine public education institutions in South Africa produced 83% of all doctorates in 2007. The top four universities, according to the report, were the University of Cape Town (11%), the University of Pretoria (13%), Stellenbosch University (12%) and the University of the Witwatersrand (11%). Together, these four universities produced 47% of all new doctorates in South Africa in 2007 (and 57% of all new doctorates produced excluding those by universities of technology). PhD production at these four universities is likely to have responded to these pressures on doctoral education. As the choice of PhD thesis format is deeply embedded within pedagogical practices and influenced by policy directives, and may be interpreted in terms of both the quantity of PhD output and the quality of the doctoral education experience, it is important to explore how the phenomenon of thesis format plays out within particular disciplinary environments. Although it is relatively easy to measure the quantity of doctoral output (provided that complete data sets are available), the quality of such output is rather more difficult to measure in any discipline (including economics with its multiple sub-disciplines). But quality does not only have to refer to the actual output produced; it can also refer to doctoral education itself – that which is experienced and enacted by the supervisors and students within the discipline. In this article we take the latter position, as only subject experts (the examiners of the PhDs) can really judge the quality of a thesis. We are not subject experts in the fields in which the included theses in our data set were produced, but as all these theses served as evidence of completed doctorates, we assumed that they must have met at least the minimum standards in terms of quality within the broader discipline of economics.

The disciplinary environment of economics forms a useful context for such analyses as PhDs in economics have an important role to play within the national doctoral scenario in which 'education, economic and management sciences, and religious studies' produced the largest share of all doctorates.[2] Economics departments produced the bulk of the doctorates within the commerce faculties in our data set. A discipline-based analysis may therefore provide interesting insights as Becher and Trowler[14] argue that academic communities are characterised by distinctive discipline-based cultures that influence their practices. Disciplines, such as economics, have ethnocentric views and thus respond differently to stimuli.

If economics doctorates have become 'commodified inputs' in a national PhD production process,[10] not only of research but also of PhD students, then pressure for increased output with the same input is to be expected. All else being equal, a reduction in the quality of doctoral education may result. Similar logic also leads to a hypothetical expectation of fewer supervisors per thesis. The ASSAf report[2] indicated that the traditional apprenticeship model of doctoral supervision, currently predominant at South African universities, might not be able to support the required increase in numbers. There are a variety of possible responses to such pressures, including (amongst others) a cohort approach[15] and the PhD by publication. A move towards the PhD by publication format may suggest a change in the purpose behind a PhD. Samuel[16] suggests that the purpose for which a PhD is undertaken has significant attributional implications. This may well extend to the lens through which quality is perceived. What the economics discipline associates with PhD thesis quality, and whether increased use of the PhD by publication form is shifting the goal posts, are addressed in this paper.

The ASSAf report, combined with a discipline-specific understanding of the study environment, thus forms the foundational rationale for this enquiry. Did the departments of economics at the top four universities respond to the call for increased doctoral production? And did the PhD by publication play an important role in a potential increase? How might an increase in quantity influence the quality of doctoral education within these academic settings? The first two quantity-orientated questions are answered by analysing a data set built from PhD theses obtained from the institutional repositories of the four universities (Phase 1). Quality of the doctoral education within these settings is explored with open-ended questionnaires to the former heads of department (Phase 2).

This paper is one of few to explore the production of PhDs in detail in one discipline. Moreover, it allows for a comparison between the economics departments at the four selected research-intensive universities in South Africa in terms of supervisory performance; this type of comparison has not been attempted before. Such discipline-specific analyses across institutions provide useful insights into the nature of the doctorate as it manifests within and across contexts. This is of interest not only to economics academics who are responsible for implementing policies and making informed pedagogical choices, but also to prospective doctoral students in economics as they make choices in terms of institutions, supervisors, programmes and formats. Compared to the ASSAf report, this paper presents and analyses data on PhDs produced in South Africa in the field of economics for an updated period.

Literature review

The increasing importance of a flexible 'knowledge economy' as an imperative for future economic growth seems to have sparked a world-wide interest in and consequent investment in the generation of knowledge in the form of doctorates. The South African government, under pressure to improve South Africa's economic growth rate, seems to expect much from the eventual dividends associated with the production of more doctorate degrees. The Carnegie Foundation and National Research Foundation of South Africa have reached broad consensus that Africa, in particular, needs 'thousands' of PhDs in order to replace an aging professoriate to serve increased student numbers in an ever more massified higher education system.[2,17]

At the same time, South Africa faces its own unique set of challenges and impediments involved in the journey to produce more doctorates. Cloete et al.[17] identified four imperatives to which doctoral production in South Africa may be subjected: growth, efficiency, transformation and quality. The present discourses concern the current status and approaches thought to be necessary to achieve each of these imperatives and, of importance, they also concern the inevitable goal to reconcile these imperatives while limiting trade-offs amongst them.

Despite the pressing need for research concerning the production of more doctorates in South Africa, research on the topic seems scant. Wolhuter[18] and Mouton et al.[19] argue that almost no quantitative studies with this focus are available. Until recently, even less research seems to have been available concerning the production of doctorates in the individual academic discipline fields, especially in economics. The only papers found were Zarenda and Rees[20] on economics education in general, and Hosking[21] on the costs and benefits of doctoral education in economics in South Africa using a case study approach.

Discipline-focused research into doctorate production in South Africa was recently advanced with an explorative quantitative study by De Jager and Frick[22]. This study showed a low rate of doctorate production in the accounting sciences, which was attributed to a chronic lack of supervisory capacity in accounting departments, underscored by few staff at the required professorial level. However, an increased research orientation in South African university accounting departments, albeit from a low base, was also found.

Conversely, it seems a plausible expectation that the production of doctorates in economics would have reacted quicker than other fields to Cloete et al.'s[17] imperatives of growth, efficiency and transformation, as the literature shows economists to have internalised rational profit-maximisation as the main device to be used for decision-making.[23,24] Fields other than economics might be slowed down in their decision-making by the consideration of factors outside the profit-maximisation framework. The same profit-maximisation logic is also used to explain the move in economics to the PhD by publication as 'the essay format allows students to avoid the effort to convert a treatise into multiple journal papers'[25]. However, we also acknowledge that economics departments do not function in isolation from the rest of the university community, and therefore institutional imperatives (which may or may not be drivers of change) may be more determining than disciplinary idiosyncrasies.

Based on the foregoing rationale, we investigated the current state of production of PhDs in economics in South Africa, with an emphasis on quantitative information. We expected to find an increase in PhD numbers, possibly at the expense of doctoral education quality. It is of course possible that efficiency gains compensate for a potential decrease in such quality. One such possible efficiency gain is the greater use of the PhD by publication format. Furthermore, we also consequently gained useful insights or views on the current research culture at economics departments, with the eventual possibility of replicating good practices in other South African economics departments.

Research approach

The research strategy was interpretive, with the objective of describing recent trends in the production of economics PhDs in South Africa. The focus was to describe what is happening, rather than how it might be explained, as the area is not well known. Once a clear description of this area became available, the focus could shift to more explanatory approaches.

An explanatory sequential mixed-method research design was employed. Inferences from the initial quantitative phase of the study (secondary data analysis) were then combined with inferences from the qualitative phase (open-ended questionnaires).

Phase 1: Quantitative data analysis

University sample selection

South Africa has 26 public higher education institutions that can be divided into three categories: 14 universities, 6 comprehensive universities and 6 universities of technology. The University of Cape Town, the University of Pretoria, Stellenbosch University and the University of the Witwatersrand are the top four universities based on PhD production and rankings. The Centre for Higher Education Transformation clusters these four universities together as high performers based on an analysis of their inputs and outputs.[26] The four universities are the leading South African research universities according to the 2015 QS World University Rankings. They are also the highest-rated South African universities according to the 2015 Times Higher Education World University Rankings, apart from the University of Pretoria, which is rated fifth after the University of KwaZulu-Natal. Doctorates produced collectively by these universities are thus expected to be representative of the doctorates produced in South Africa, even though the sample obvious`ly does not include all the doctorates produced in the country.

Data and method

Institutional repositories of doctoral theses at the four universities (at Stellenbosch University the terminology is doctoral dissertation and master's thesis) were used to access doctoral theses for all commerce faculties (including business schools) for the period from 2008 to 2014. This choice of time period was motivated by the fact that Stellenbosch University began uploading all doctoral theses to their public repository only in 2008. The data reported in the ASSAf report[2], which at present is arguably the most comprehensive published report on the status of the PhD in South Africa, only included the period from 2000 to 2007. This paper should therefore provide a focused analysis of subsequent trends. The list of theses downloaded was reconciled to the graduation records of the four universities for completeness.

The following data were obtained from each doctoral thesis:

- the year of graduation;
- the name of the student;
- the title of the thesis;
- the academic department;
- the university affiliation;
- the names and affiliations of the supervisors;
- the type of doctoral degree (PhD/DCom);
- the number of words in the title;
- the total number of pages (as included in the pdf file);
- the number of chapters; and
- whether the doctorate was by monograph or by publication.

Determination of 'by publication' was based on an analysis of the title, the index, the abstract and chapter one of each thesis. A 'by publication' doctorate is structured as a number of papers with a global introduction and conclusion. By implication, the literature review and research approach are not separate chapters as in a monograph but are rather spread across the different papers.

It must be emphasised that only publically available information was gathered. Universities are not willing to share additional data, such as the number of years a student was registered or the personal information of students, due to privacy concerns related to the *Protection of Personal Information Act*.[27]

Despite several difficulties encountered in collecting the data, including missing records and limited online search functionality, repository and graduation records were ultimately reconciled. Economics theses were identified and extracted from the larger combined commerce faculty collection based on departmental affiliation. The data gathered enabled the analysis of quantity trends in PhD production, as well as the identification of some characteristics of those PhDs. Comparisons between the economics group of PhDs and the rest of the commerce faculty group of PhDs were possible. Comparisons between the four universities were also possible.

Phase 2: Questionnaires

Only tentative inferences about the quality of doctoral education are possible with the quantitative data. To deepen our understanding of what is happening in the production of economics PhDs in South Africa we sent open-ended questionnaires to the former heads of the four departments included in this study; the respondents were thus heads of department (HODs) for most of the 2008 to 2014 period when the theses included in the study were produced. All four purposively selected respondents responded positively to our request to participate. Ethics clearance was obtained from the University of Cape Town's Commerce Faculty Ethics in Research Committee before sending out the questionnaires (reference number: 0210201501). Informed consent was obtained from each respondent. The responses to the questionnaires enabled a triangulation of inferences based on different sources of information. The questions are available in the Appendix. The responses were analysed by means of conventional content analysis, as described by Hsieh and Shannon[28].

Results

Quantity of PhDs produced

Table 1 details the number of doctorates produced per university for each of the 7 years under review, in economics and within each commerce faculty. Table 1 also indicates the percentage of the economics PhDs that were by publication.

Table 1 shows an increase in the total number of economics PhD graduates who graduated annually over the studied period. At the same time the use of the PhD by publication format grew: in 2013, almost 70% of all PhDs in economics were in this format. The PhD by publication format has become entrenched and seems to be growing rapidly in uptake as most economics PhDs in 2013 and 2014 were in this format in all of the four universities studied. This finding accords with international experience as 65% of economics PhDs in the USA for the period 2001 to 2002 were in this format.[25]

Economics PhD graduates formed a significant part of the total commerce faculty cohort of doctoral graduates, indicating the relative importance of economics as a producer of doctorates within the larger commerce field. The percentage of the faculty doctorates from economics was relatively stable between the beginning and end years of observation, implying that economics PhD output grew at about the same pace as faculty doctorate output.

Of the four universities, the University of Cape Town produced the highest number of economics PhDs. The numbers graduated seem to have grown, albeit at a slower pace than the numbers graduated in the faculty as a whole as indicated by the percentage of the faculty's doctorates from economics indicator.

Table 1: Total PhDs in economics produced in the years 2008 to 2014 at the Universities of Cape Town, Pretoria, Stellenbosch and the Witwatersrand

Year	Total				Cape Town				Pretoria				Stellenbosch				Witwatersrand			
	Economics	% Economics by publication	Faculty	% Faculty by economics	Economics	% Economics by publication	Faculty	% Faculty by economics	Economics	% Economics by publication	Faculty	% Faculty by economics	Economics	% Economics by publication	Faculty	% Faculty by economics	Economics	% Economics by publication	Faculty	% Faculty by economics
2008	**12**	25%	**44**	27%	4	50%	7	57%	5	0%	24	21%	1	0%	11	9%	2	50%	2	100%
2009	**9**	44%	**44**	20%	7	43%	19	37%	2	50%	21	10%	0	NA	4	0%	0	NA	0	NA
2010	**11**	36%	**42**	26%	7	57%	17	41%	0	NA	7	0%	0	NA	12	0%	4	0%	6	67%
2011	**26**	31%	**69**	38%	8	63%	18	44%	9	0%	25	36%	3	33%	15	20%	6	33%	11	55%
2012	**21**	29%	**64**	33%	10	20%	26	38%	3	67%	15	20%	5	20%	16	31%	3	33%	7	43%
2013	**15**	67%	**55**	27%	7	71%	14	50%	5	80%	28	18%	1	0%	7	14%	2	50%	6	33%
2014	**20**	55%	**73**	27%	6	67%	16	38%	8	50%	29	28%	6	50%	16	38%	0	NA	12	0%
TOTAL	**114**	**40%**	**391**	**29%**	**49**	**51%**	**117**	**42%**	**32**	**34%**	**149**	**21%**	**16**	**31%**	**81**	**20%**	**17**	**29%**	**44**	**39%**

The data were sourced from the public repositories of these four universities. PhDs by publication were identified based on self-classification by the student or the format of the thesis.

However, the School of Economics at the University of Cape Town was a more significant contributor to the doctoral output of the commerce faculty (42% on average) than the other three departments of economics sampled (29% on average for the four universities combined). A high and increasing percentage of the economics PhDs at Cape Town were in the PhD by publication format.

The second-highest number of economics PhDs was produced at the University of Pretoria. Of the four universities, the University of Pretoria's commerce faculty produced the most commerce doctorates. However, the percentage of economics PhDs in the commerce faculty was the second lowest of the four universities at 21%, but stable. The number of economics PhDs at the University of Pretoria grew during the study period. A lower average percentage of the PhDs was in the PhD by publication format compared with the four universities combined; however, this trend seems to be changing when looking at the data for the period 2012 to 2014.

The commerce faculty at Stellenbosch University produced the second-lowest number of doctorates overall. That faculty also produced the lowest overall number of economics PhDs. The contribution from economics PhDs to faculty doctorates was the lowest at Stellenbosch University of the four universities on average. However, the number of economics PhDs grew steadily over the period under review. A fast-growing percentage of the economics PhDs was in the PhD by publication format at Stellenbosch University.

The University of the Witwatersrand produced the second-lowest number of economics PhDs in total. Also, the total number of doctorates at faculty level was the lowest of the four universities. A notable, but declining, contribution came from economics. A firm percentage (29%) of the economics PhDs at the University of the Witwatersrand were by publication.

The PhDs in economics produced at the four universities were concentrated amongst particular supervisors. Supervisor A at the University of Cape Town produced the most (10) PhDs and Supervisor B at the University of Pretoria produced the second-most (nine) PhDs during the period. Supervisor C at Stellenbosch University, Supervisor D at the University of Pretoria and Supervisor E at the University of the Witwatersrand produced six PhDs each during the period. This concentration raises important quality-related questions about the doctoral education practices within these departments, as such supervisors might have been overloaded. Given the aging of productive supervisors, and the relatively low percentage of doctorated academic staff within the South African university system in general, this result is not necessarily surprising – but it does bring into question how more and younger supervisors are supported whilst not compromising on the quality of doctoral education.

This perspective on quantity should not be divorced from other aspects of the doctorates under consideration, such as the quality of doctoral education. Although the selected attributes presented below cannot be seen as accurate indicators of the quality of doctoral theses, they do provide an interesting perspective on trends in economics doctoral thesis inputs (doctoral education) and outputs (theses).

Selected attributes of the economics PhDs

Table 2 details the evolution of the average number of supervisors per economics PhD and commerce faculty doctorate per year. Three length indicators are detailed per economics PhD and commerce faculty doctorate per year and per university: average number of pages, average number of chapters and average number of words in the title. Table 3 details the 20 most common words used in the titles of the economics PhDs versus the 20 most common words used in the titles of the other commerce faculty doctorates.

Table 2 indicates that the overall mean number of supervisors for the economics PhDs was almost the same as the overall mean number of supervisors for the rest of the commerce faculty doctorates. However, the overall mean number of supervisors differed strongly between universities, both for the economics group and the rest of the commerce faculty group. The Universities of Cape Town and the Witwatersrand clustered together with a lower number of supervisors per doctorate compared to the number for Pretoria and Stellenbosch universities. On a per-year level, it seemed as if the mean number of supervisors per economics PhD decreased, whilst for the rest of the commerce faculty doctorates it stayed stable or was even slightly increasing.

In terms of page length of the thesis, the economics PhD was significantly shorter than those from the rest of the commerce faculty for each of the years under review as well as overall. The average page length was significantly different among universities for the economics group but not for the rest of the commerce faculty group. The economics PhDs at the University of Pretoria were much shorter on average than those at the other three universities. On a per-year level the average number of pages per economics PhD decreased, whilst for those from the rest of the commerce faculty, the number of pages stayed stable. The average number of chapters in economics PhDs was significantly shorter than those in PhDs from the rest of the commerce faculty, overall and for four of the years under review. The average number of chapters was not significantly different between the universities for the economics group and for the rest of the commerce faculty group. On a per-year level, it seems as if, for the last two years under review, the number of chapters per PhD decreased in economics, whereas it remained stable for PhDs from the rest of the commerce faculty. In terms of title length, the economics PhD group produced significantly shorter titles than the rest of the commerce faculty group overall and for three of the years under review. The average length of PhD title at Stellenbosch University was significantly longer than those at the other universities for the economics PhD group. On a per-year level, the average title length of economics PhDs increased, whereas it remained stable for PhDs from the rest of the commerce faculty.

Table 3 indicates that both economics PhDs and the rest of the commerce faculty doctorates were mostly contextually situated to South Africa and Africa. There was a 35% overlap between the common words used in the titles of the economics PhDs and the common words used in the titles of the rest of the commerce doctorates. The word 'essays' was used in 16 out of the 114 economics titles (for example, 'Three essays on…'). Such a title was interpreted to indicate the PhD by publication format.

This analysis of selected attributes of the doctorates shows that the increase in the quantity of economics PhDs was accompanied by a decrease in the number of supervisors, as well as decreases in the number of pages and chapters. However, this was not the case for the rest of the commerce faculty. These results, which could be interpreted as negative for the quality of doctoral education, could also be interpreted as indicating an improvement in the quality of doctoral education: that is, the number of supervisors in economics might have reduced because of a new cohort approach to PhD supervision or the number of pages (and chapters) might have reduced because economics PhD students became more concise in their writing. Conciseness is a characteristic of high-quality writing.[29,30] Although the quantitative data cannot be used to distinguish between these opposing possibilities, this quandary does warrant further inquiry.

Multivariate analysis

Our overall data set contained both economics PhDs and doctorates from the rest of the commerce faculties, which afforded us the opportunity to investigate any factors distinguishing the economics PhDs from the rest of the commerce faculty doctorates in a multivariate environment. We made use of a logit regression as the dependent variable was binary. Possible explanatory variables identified were: the university; the year of

Table 2: Selected attributes of the 114 economics PhDs produced versus the 277 PhDs produced by the rest of the commerce faculty

	Length indicators							
Year	**Number of supervisors**		**Number of pages**		**Number of chapters**		**Number of words in title**	
	Economics combined	**Faculty†**	**Economics combined**	**Faculty†**	**Economics combined**	**Faculty†**	**Economics combined**	**Faculty†**
2008	1.417	1.281	225.750***	323.250	7.000	7.719	10.667*	13.594
2009	1.333	1.257	227.222***	325.314	6.778	7.086	13.778	16.057
2010	1.273	1.258	180.727***	311.645	5.545***	7.742	10.364**	13.581
2011	1.385	1.279	190.577***	324.738	6.385**	7.500	10.769**	12.814
2012	1.095	1.209	239.900**	298.310	6.800	7.024	13.762	13.628
2013	1.267	1.475	226.286***	322.079	5.857***	7.632	10.133	10.275
2014	1.300	1.321	207.158***	322.660	6.000***	7.547	11.700	12.679
Total	**1.289**	**1.300**	**212.604*** **	**318.311**	**6.342*** **	**7.454**	**11.588*** **	**13.134**
Cape Town	1.082	1.162	212.417	296.908	6.292	7.292	11.653	15.206
Pretoria	1.656	1.359	172.667	332.188	6.333	7.479	10.125	12.111
Stellenbosch	1.375	1.385	247.625	321.308	6.750	7.738	15.188	12.877
Witwatersrand	1.118	1.185	250.647	301.311	6.118	7.038	10.765	12.963
Different: Anova F-test	0.000	0.018	0.001	0.101	0.750	0.363	0.003	0.002
Different: Welch F-test	0.000	0.013	0.003	0.107	0.812	0.249	0.041	0.022

The data were sourced from the public repositories of the four universities.

†Excluding economics.

Statistically significant differences are shown between the mean for the economics PhDs versus the mean for the remainder of the commerce faculty doctorates at the 10% (*), 5% (**) and 1% (***) levels. In the blocked section at the bottom of the table the overall average per variable is decomposed into an average of that variable per university. The Anova F-test and Welch F-test indicate whether these university differences are statistically significantly different from each other.

Table 3: Word frequencies (top 20) in the titles of doctoral theses at the Universities of Cape Town, Pretoria, Stellenbosch and the Witwatersrand

Economics			Rest of faculty		
Word	**Count**	**Weighted %**	**Word**	**Count**	**Weighted %**
africa	**42**	**4.34**	**south**	**104**	**3.96**
south	**41**	**4.24**	**africa**	**73**	**2.78**
economic	26	2.55	**african**	**56**	**2.13**
modelling	**19**	**1.96**	**development**	**67**	**2.12**
african	**18**	**1.86**	**model**	**41**	**1.56**
policy	**18**	**1.79**	study	50	1.39
essays	16	1.65	management	39	1.37
market	**16**	**1.65**	evaluation	45	1.36
value	17	1.47	governance	57	1.35
financial	14	1.45	role	34	0.97
monetary	13	1.34	public	32	0.97
outcomes	16	1.29	business	26	0.88
growth	16	1.24	case	25	0.86
poverty	11	1.14	**market**	**23**	**0.84**
analysis	10	1.03	systems	35	0.81
evidence	10	1.03	performance	25	0.78
education	13	0.93	based	27	0.76
countries	9	0.88	effects	30	0.74
development	**16**	**0.88**	work	39	0.71
dynamics	8	0.83	**policy**	**19**	**0.67**

The data were sourced from the public repositories of the four universities. Words in bold indicate overlap between the two groups.

graduation (as economics PhDs in total seemed to be increasing, the variable was defined as a dummy variable equal to one if the graduation year was 2012, 2013 or 2014); the format (by publication or not); the supervisor (the variable was defined as a dummy variable equal to one if the supervisor was the supervisor of more than five doctorates in our data set); the number of supervisors; the number of pages; the number of chapters; and the number of words in the title. The full model was estimated and any insignificant variables were dropped from it, starting with the most insignificant variable. Table 4 details the resulting final model.

Table 4 demonstrates that in a multivariate environment there was not a statistically significant increase in economics PhDs compared to the rest of commerce faculty doctorates as the dummy variable was dropped from the final regression. The most productive supervisor dummy variable was also dropped from the final regression and so the mere fact that a student was supervised by a productive supervisor did not uniquely identify the doctorate as one from economics. The number of supervisors was also dropped from the final regression. Of the length variables, only the number of pages of the thesis came through as a significant predictor of an economics PhD in the final model.

The final model can be interpreted as follows. The base case is an economics PhD at the University of Pretoria. If it is known that the thesis format is by publication then the probability of the doctorate being an economics PhD increases by 19.4%. If it is known that the student is not from the University of Pretoria and rather from Stellenbosch University then the probability of the doctorate being an economics PhD is unchanged from the probability of the University of Pretoria. If it is known that the student is not from the University of Pretoria and rather from the University of Cape Town then the probability of the doctorate being an economics PhD increases by 12.5%. If it is known that the student is not from the University of Pretoria and rather from the University of the Witwatersrand then the probability of the doctorate being an economics PhD increases by 18.2%. Finally, a shorter thesis is more likely to be an economics PhD; a one-page increase in length (compared to the mean page length of 288 pages) was associated with a 0.3% decrease in the probability that a doctorate is in economics.

The quantitative data analysis confirms that the quantity of economics PhDs is increasing, that the format is changing away from the monograph PhD to the PhD by publication and that the form of the economics PhDs is showing trends that are different from the rest of the commerce faculty. Experts within the discipline were approached to gain a perspective on the quality of these changes.

Table 4: Final logit regression model that distinguishes between economics PhDs and PhDs produced in the rest of the commerce faculty

Regressand: Economics=1			
Regressor	**Coefficient**	**Z-statistic**	**Marginal effects**
Constant	2.144	4.119***	
By publication=1	0.987	2.961***	0.194103985
Cape Town=1	0.636	2.113**	0.125076124
Witwatersrand=1	0.925	2.200**	0.181911029
Page length	-0.014	-7.094***	-0.002753248
Observation with Economics=0	273		
Observation with Economics=1	111		
N	384		
McFadden R-squared	0.272		
Probability (LR statistic)	0.000		

The data were sourced from the public repositories of the four universities.

Marginal effects were derived from the model coefficients and the means of the variables. The base case was an economics PhD at the University of Pretoria.

Statistically significant coefficient at the 5% (**) and 1% (***) levels.

Questionnaires and discussion

Former HODs, who served for most of the period of the study (2008–2014) in the four economics departments, were asked to reflect on the PhD production processes in their respective departments.

The HOD respondents identified building a research culture and developing departmental research capacity as the main drivers behind the increased PhD production noted in the quantitative data. The HOD respondents were divided on whether their respective departments had mature research cultures that could support increased PhD production. Most respondents were convinced that this was the case, but even the one respondent who did not share this view indicated that there had been substantive improvements over the past 5 years. The main factors that contributed towards the establishment of a mature research culture within departments were: attracting international and esteemed scholars (with PhDs) onto the permanent staff complement; providing staff members with international exposure (through research seminars and/or overseas study opportunities); prioritising and incentivising research (rather than private consulting) as a core academic function and responsibility;

developing strong research focal areas (and even formalised research units) around which staff members and postgraduate students collaborate and co-publish; and building strong and research-focused postgraduate programmes. Thus a department with a research-focused ethos and highly qualified staff members who could network, facilitated the development of a mature research culture.

The noted increase in quantity also had a perceived positive influence on research quality in general, as the following responses indicate:

> *We want to be a world-class Economics Department. …In the beginning it requires much input without much output, but after about four to five years the investment starts paying off as the PhD students start to graduate. …We found that when we were hiring a substantial number of staff in 2009/2010, and we were looking in the international pool, having a large PhD programme was very beneficial, and attracted a much better cohort of candidates than had been the case otherwise.* [Respondent 1 (R1)]

> *For us, this is about undertaking quality research, and training people to do the same. We like to do research, there are tons of interesting questions, but we cannot get to them all. So, PhD students can help us address a few more questions than we would be able to take on ourselves.* [R3]

The quality of PhDs produced, in particular, was also perceived by the HODs to be better than before. The indicators they used to qualify the improved quality output included the improved PhD student selection criteria, which could (in part) be related to selecting students whose research interests and background matched that of a department's research consortia. One particular HOD referred to an increase in PhD student numbers related to the establishment of such a research niche area as creating a catchment area for excellent students, with the added spin-off of creating a critical mass of students who could support each other.

In the quantitative data, we noted a marked increase in the proportion of PhDs by publication, and therefore enquired why this might be the case (Question 5). One of the HODs attributed this increase to international trends:

> *This is a nod to the international market. It has been the norm in the US and Europe, where many of us trained, and, therefore, we have implemented it with many of our students. There are still students who are not 'built' to work that way, so we still have the more traditional 'book' PhD now and then.* [R3]

Interesting here is the use of 'market', which supports our earlier notion that economics departments may be characterised by a distinctive discipline-based culture that influences their practices.

Other possible reasons that HODs gave for an increase in the PhD by publication format within the departments studied were mainly attributed to the incentives that supervisors received for publications and the opportunities for wider dissemination of the work:

> *At the end of the day, if it is not published, it is not done, or so the adage goes.* [R1]

> *It seems like a more direct way to get the benefit of joint publications for supervisors.* [R2]

> *To encourage and facilitate the publication and wider dissemination of the research results.* [R4]

Whilst the HODs did not refer to this consideration directly, the financial incentive for publication through the subsidy system (peculiar to the South African higher education system) may account, partly, for the increase in PhDs by publication. The respondents did not see any direct financial benefit for the department from producing a PhD thesis alone:

> *Not directly, but indirectly. We did not directly receive subsidy money.* [R1]

> *The department has gradually benefitted from this, and the incentive to supervise students has now improved, but I do not know if this has made members of staff more willing to take on supervision.* [R2]

> *Not in any way that we can see.* [R3]

> *Indirectly, yes.* [R4]

The HODs were also asked (Question 2) to comment on the main institutional factors that, in their view, inhibited the production of larger numbers of PhD graduates from their departments. The lack of supervisory capacity seemed to be the main hampering factor reported in the case of the four departments studied:

> *Supervisory capacity. We have quite a number of professors, but even so the distribution of the supervision load is fairly uneven. Some supervisors are overloaded.* [R1]

> *The main institutional factor was the lack of a tradition of producing PhDs, the extent to which most of the top students went overseas to get their PhDs and a lack of willingness/capacity amongst some staff members to supervise PhDs.* [R2]

It seems that, once a department is able to establish a vibrant research culture, with enough doctorated staff members to sustain the research impetus, a tipping point is reached at which an increase in PhD production can be achieved and sustained.

Funding was also noted as a key deterrent for PhD studies, as students would not be able to study full-time without such support. Part-time studies were not seen as a viable option for many students, especially if they were already in positions of full-time employment (where a PhD in economics would not necessarily benefit them professionally).

Thus, the qualitative information shows that an increase in the quantity of economics PhDs was not accompanied by a concurrent decrease in the quality in doctoral education as much as it was accompanied by a change in departmental research management. Two changes stand out: the establishment of a research culture and the move to the PhD by publication. These management changes will be briefly discussed further.

In the four departments included in the study, national and international calls for increased PhD production coincided with recognition of the importance of establishing a research culture and prioritising research as an academic activity within economics departments. Wills et al.[31] underscore the importance of fostering a research culture, securing adequate time and funding for research, and having staff with PhDs who are productive in publishing early in their careers as essential elements for research productivity of accounting academics. Kyvik and Aksnes[32] found that better qualified academic staff, increased research collaboration, improved research conditions, and incentive and reward systems all helped to explain an increase in research productivity at Norwegian universities. These elements were also highlighted by the HODs in our study as being drivers of both quantity of PhD output and quality in doctoral education in their economics departments.

The use of the PhD by publication format was indicated to be of special relevance to economics as a discipline, by both the literature review and the logit regression results. Callaghan[33] recommends this alignment of publication practices and higher degree requirements – thus the PhD by publication – as a way of building human capital to boost higher education research productivity. However, the use of the PhD by publication format addresses mainly the interests of the department, the supervisor and the doctoral candidate who aims to pursue an academic career, rather than the needs of those who come from and/or re-enter non-academic jobs.[25] Moreover, the PhD by publication fundamentally requires expert supervisory support[34], preferably given by supervisors engaged in publication themselves[35].

If one takes into account the relatively low percentage of staff with doctorates at South African universities (including economics departments, as highlighted by the HODs in this study), it would seem that attempts to find suitably qualified supervisors to guide students following this format might be problematic if it becomes a requirement to bolster doctoral production. Despite these cautionary considerations, Stock and Siegfried[25] found that the PhD by publication could significantly decrease the time needed for completion of the doctorate. So while the PhD by publication might certainly hold advantages in terms of PhD production and building scholarship within economics in South Africa, it should not be seen as an 'easy way out'.[36] As is the case in Australia[37] and New Zealand,[38] South Africa also lacks a nationally consistent guideline on the quality and quantity parameters that define this PhD format, which leaves much room for (mis-) interpretation by academic departments, supervisors and examiners.

Conclusion

We set out to explore if the departments of economics at the top four universities, as defined in this paper, responded to the call for increased doctoral production. If so, was there any increase at the expense of quality of doctoral education and did the PhD by publication play a role in any of these shifts?

The four economics departments and the commerce faculties at these universities are producing substantially more PhDs – thus quantity is increasing. The indicators gathered show the economics PhD group and the trends in the form of that groups' PhDs to be different from the rest of the commerce faculty in a way that can be interpreted as either positive or negative in terms of quality. This ambiguity necessitated the gathering of additional survey data. Former HODs of the economics departments supplied a plausible argument as to why the quality of doctoral education in these departments had improved: management changes to the process of economics PhD production. Such changes include the move to the PhD by publication, increased attention to ensuring the quality of students allowed into the PhD programmes, facilitation of full-time doctoral studies through funding arrangements, and the appointment of international faculty with a research orientation. We thus infer that increases in quantity did not necessarily have an adverse influence on the quality of doctoral education.

Our quantitative data, and the absence of a generally accepted quantitative model that measures PhD thesis quality, did not allow us to make inferences about both quantity and quality of theses. This limitation to our study, at the same time, encourages researchers to develop such a model with quantitative indicators for PhD thesis quality.

Acknowledgements

We gratefully acknowledge the financial support of the Research Office at the University of Cape Town.

Authors' contributions

P.d.J. was the lead author. P.d.J., L.F. and P.v.d.S. all contributed conceptually. P.d.J. collected the data and performed the quantitative analysis. L.F. performed the qualitative analysis. All authors contributed towards the final write-up.

References

1. Summers L. No free lunches but plenty of cheap ones. Financial Times. 2016 February 7. Available from: http://www.ft.com/cms/s/2/43f9002e-cc04-11e5-a8ef-ea66e967dd44.html#axzz4GBLoaGZj
2. Academy of Science of South Africa (ASSAf). The PhD study consensus report. Pretoria: ASSAf; 2010. Available from: http://www.assaf.org.za/files/2010/11/40696-Boldesign-PHD-small.pdf
3. National Planning Commission, Office of the Presidency. National development plan [document on the Internet]. c2012 [cited 2016 Feb 29]. Available from https://nationalplanningcommission.files.wordpress.com/2015/02/ndp-2030-our-future-make-it-work_0.pdf
4. Higher Education South Africa. South African higher education in the 20th year of democracy: Context, achievements and key challenges [document on the Internet]. c2014 [cited 2016 Feb 29]. Available from http://www.hesa.org/hesa-presentation-portfolio-committee-higher-education-and-training
5. Mouton J. Post-graduate studies in South Africa: Myths, misconceptions and challenges. S Afr J Higher Educ. 2007;21(8):1078–1090. http://dx.doi.org/10.4314/sajhe.v21i8.25761
6. Teferra D. Manufacturing and exporting excellence and 'mediocrity': Doctoral education in South Africa. S Afr J Higher Educ. 2015;29(5):8-19.
7. Thaver B, Holtman L. A response to Yusef Waghid's leading article. S Afr J Higher Educ. 2015;29(5):20–28.
8. Waghid Y. Are doctoral studies in South African higher education being put at risk? S Afr J Higher Educ. 2015;29(5):1–7.
9. Parker LD. From privatised to hybrid corporatised higher education: A global financial management discourse. Financ Account Manage. 2012;28(3):247–268. http://dx.doi.org/10.1111/j.1468-0408.2012.00544.x
10. Guthrie J, Parker D. The global accounting academic: What counts! Account Audit Accountability J. 2014;27(1):2–14. http://dx.doi.org/10.1108/AAAJ-10-2013-1504
11. Badley G. Publish and be doctor-rated: The PhD by published work. Qual Assurance Educ. 2014;17(4):331–342. http://dx.doi.org/10.1108/09684880910992313
12. Boud D, Lee A. Changing practices of doctoral education. Oxford: Routledge; 2009.
13. Louw J, Miller J. A literature review on models of the PhD [document on the Internet]. c2015 [cited 2015 Aug 30]. Available from http://chet.org.za/papers/literature-review-models-phd
14. Becher T, Trowler PR. Academic tribes and territories. Buckingham: Society for Research in Higher Education; 2001.
15. Samuel M, Vithal R. Emergent frameworks of research teaching and learning in a cohort-based doctoral programme. Perspect Educ. 2011;29(3):76–87. http://hdl.handle.net/10413/11029
16. Samuel M. Values and purposes of a PhD: Comparative responses from South Africa and Mauritius. Higher Education Forum. 2016;13:19. http://doi.org/10.15027/39937
17. Cloete N, Mouton J, Sheppard C. Doctoral education in South Africa. Cape Town: African Minds; 2015.
18. Wolhuter C. Research on doctoral education in South Africa against the silhouette of its meteoric rise in international higher education research. Perspect Educ. 2011;29(3):126–138. http://hdl.handle.net/10394/7776
19. Mouton J, Boshoff N, James M. A survey of doctoral supervisors in South Africa. S Afr J Higher Educ. 2015;29(2):1–22.
20. Zarenda H, Rees D. Economics education in South Africa output analysis. S Afr J Eco. 1984;52(2):190–195. http://dx.doi.org/10.1111/j.1813-6982.1984.tb00830.x
21. Hosking SG. The costs and benefits of doctoral education in economics in South Africa. S Afr J Econ. 1997;65(2):112–117. http://dx.doi.org/10.1111/j.1813-6982.1997.tb01362.x
22. De Jager P, Frick L. Accounting doctorates produced in South Africa 2008-2014. Meditari Account Res. 2016:24(3):438–457. http://dx.doi.org/10.1108/MEDAR-06-2015-0033.
23. Cipriani GP, Lubian D, Zago A. Natural born economists. J Econ Psychol. 2009;30(3):455–468. http://dx.doi.org/10.1016/j.joep.2008.10.001
24. Wang L, Malhotra D, Murnighan JK. Economics education and greed. Academy of Management Learning & Education. 2011;10(12):643–660. http://dx.doi.org/10.5465/amle.2009.0185
25. Stock W, Siegfried J. One essay on dissertation formats in economics. Am Econ Rev. 2013;103(3):648–653. http://dx.doi.org/10.1257/aer.103.3.648
26. Bunting I, Sheppard C, Cloete N, Balding L. Performance indicators in South African higher education [document on the Internet]. c2010 [cited 2015 Aug 30]. Available from http://www.chet.org.za/books/performance-indicators-south-african-higher-education-2000%E2%80%932008

27. Government Gazette: Republic of South Africa. Protection of Personal Information Act 4 of 2013. Available from: http://www.justice.gov.za/legislation/acts/2013-004.pdf

28. Hsieh HF, Shannon S. Three approaches to qualitative content analysis. Qual Health Res. 2005;15(9):1277–1288. http://dx.doi.org/10.1177/1049732305276687

29. Ortrun Z, Fletcher M. The quality of an action research thesis in the social sciences. Qual Assurance Educ. 2007;15(4):413–436.

30. Letchford A, Moat HS, Preis T. The advantage of short paper titles. R Soc Open Sci. 2015;2, 6 pages. http://dx.doi.org/10.1098/rsos.150266

31. Wills D, Ridely G, Mitev H. Research productivity of accounting academics in changing and challenging times. J Account Organ Change. 2013;9(1):4–25. http://dx.doi.org/10.1108/18325911311307186

32. Kyvik S, Aksnes DW. Explaining the increase in publication productivity among academic staff: A generational perspective. Stud High Educ. 2015;40(8):1438–1453. http://dx.doi.org/10.1080/03075079.2015.1060711

33. Callaghan CW. Higher education research productivity: The influences of different forms of human capital. S Afr J Higher Educ. 2015;29(5):85–105.

34. Robins L, Kanowski P. PhD by publication: A student's perspective. J Res Pract. 2008;4(2):1-20. http://jrp.icaap.org/index.php/jrp/article/view/136/154

35. Pare A. Slow the presses: Concerns about premature publication. In: Aitchison C, Kamler, B, Lee A, editors. Publishing pedagogies for the doctorate and beyond. Oxon: Routledge; 2010. p. 30–46.

36. Niven P, Grant C. PhDs by publications: An 'easy way out'? Teach Higher Educ. 2012;17(1):105–111. http://dx.doi.org/10.1080/13562517.2012.640086

37. Jackson D. Completing a PhD by publication: A review of Australian policy and implications for practice. Higher Educ Res Dev. 2013;32(3):355–368. http://dx.doi.org/10.1080/07294360.2012.692666

38. Sharmini S, Spronken-Smith R, Golding C, Harland T. Assessing the doctoral thesis when it includes published work. Assess Eval Higher Educ. 2015;40(1):89–102. http://dx.doi.org/10.1080/02602938.2014.888535

Appendix: Questionnaire sent to former heads of economics departments at four research-intensive universities in South Africa

1. From a departmental perspective what is the primary motivation to graduate more doctoral students?
2. From a departmental perspective what is the main institutional factor that inhibits the graduation of more doctorates?
3. Did the quality of the doctorates produced 2008 to 2014 decrease, increase or stay the same? What indicator are you using to answer?
4. Did your department benefit from increased funding from your university when you graduated a doctoral student?
5. What motivated the increased use of the 'PhD by publication' in your department?
6. Do you think that your department has a mature research culture?
7. If so, what was the main factor that contributed towards the establishment of a mature research culture in your department?

The enduring and spatial nature of the enterprise richness of South African towns

AUTHOR:
Daan F. 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: 27 June 2016
Revised: 06 Sep. 2016
Accepted: 25 Sep. 2016

KEYWORDS:
entrepreneurial spaces; enterprise dynamics; enterprise ecology; regional economics; business evolution

HOW TO CITE:
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. http://dx.doi.org/10.17159/sajs.2017/20160190

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

Enterprise richness (measured by the number of enterprise types) showed a statistically significant log-log relationship (or power law) with the total number of enterprises in (1) towns in different regions of South Africa and (2) towns in the same region but seven decades apart. Entrepreneurial space in towns develops or disappears in a regular way as towns grow or regress, which is further proof of orderliness in the enterprise dynamics of South African towns. The power laws are very similar to one another, which was powerfully illustrated by the fact that one relationship extracted from seven-decade-old information could accurately predict the enterprise richness of modern towns in South Africa. The enterprise richness power law of towns in South Africa extends over space and time. Recent reviews of research on small towns and local economic development in South Africa have ignored the orderliness detected in their enterprise structures. Islands have provided laboratories for the study of natural evolution and the MacArthur–Wilson Species Equilibrium Model based on island biogeography was a main contributor to progress in ecology. Research on regional economic geography in South Africa should move beyond the merely descriptive/narrative to more quantified research. In considering the lack of employment and poverty in South Africa, the National Development Plan suggests that towns and rural areas are important cogs in efforts to overcome these problems. Development plans that are out of sync with the observed regularities are perhaps bound to fail.

Significance:

- Prediction of enterprise composition and dynamics in South African towns.

Introduction

There are similarities between economic wealth and biological wealth because they are thermodynamically the same sort of phenomena. Beinhocker[1] stated that both are systems of low entropy and patterns of order that have evolved over time under constraint of fitness functions. Like a living organism, each individual enterprise is in constant competition for survival and only the fittest survive. The similarities between living organisms and enterprises offer the opportunity to apply lessons from natural ecology to the study of enterprise dynamics. For instance, South African towns have been likened to enterprise ecosystems[2] or to 'islands in a sea of farms'[3].

Most people would intuitively accept that larger towns have a higher abundance of different types of enterprises than smaller towns. If so, can the phenomenon be quantified and predicted? These questions led to a study of a group of 134 South African towns.[4] It was found that the abundance of enterprise types is a function of the total enterprise numbers of the towns and follows a power law with an exponent of 0.7164. (Ball[5] remarked that there is no hidden Hobbesian significance in the word 'power' – it is just a mathematical term. The term power law is used as a mathematical term in this contribution.) Therefore, for every doubling of the number of enterprises in a town, the abundance of enterprise types will increase by 71.6%. The power law confers quantitative predictability – a utility recently employed to predict impacts of potential shale gas exploitation on the enterprise structures of Karoo towns.[6]

The term enterprise richness was proposed to describe the abundance of enterprise types in towns[4]. It is analogous to the term species richness used in natural ecology, which is a measure of the variety of species, i.e. the number of species, in a given area or in a given sample.[7] Species richness is a component of the MacArthur–Wilson Species Equilibrium Model, which marked 'the coming of age of ecological science as a discipline with a theoretical/conceptual base'[8]. Simply put, this model proposed that the rate at which new species immigrate to an island is balanced by the rate at which species are eliminated from the island, hence a species equilibrium is reached. By the same token, one could postulate that if towns were 'islands in a sea of farms'[3], the rate by which new enterprise types appear in the business structure of a town should be balanced by the rate at which existing enterprise types disappear from the business structure of the town, resulting in an equilibrium of enterprise types, i.e. an enterprise richness equilibrium. Results obtained so far[4] seem to support this contention.

Further investigation of enterprise richness as a potentially useful component of attempts to improve the quantitative understanding of enterprise dynamics in South African towns, raised three questions: (1) Is it important to study the enterprise dynamics of South African towns? (2) Are there differences in the enterprise richness–total enterprises relationships of different regions of South Africa? (3) Does the enterprise richness–total enterprises relationship change over time?

Purpose of this study

The prime purpose of this study was to address the three questions outlined above by (1) overviewing research on small towns in South Africa, (2) examining the enterprise richness power laws of three South African regions – the Free State[9], the Karoo[6] and the Gouritz Cluster Biosphere Reserve (GCBR)[10] – and (3) testing the durability of the enterprise richness power law of a single region (the Free State) in two different periods seen decades apart. More context – which sheds additional light on the possibility that an enterprise richness equilibrium model might be attainable – is provided before the analyses are presented.

Research on small towns in South Africa

The National Planning Commission[11] noted that, at present, most South Africans live in a complex network of towns and cities, which generates about 85% of all economic activity. A review of recent themes in the investigation of small towns in South Africa argued that a narrow set of issues, mainly concerning economic expansion and change, have formed the key vantage point from which to view small town South Africa.[12] Depopulation and small town decline[13] and migration of individuals and households to places that are better able to provide livelihoods[14] set the scene against which the National Development Plan 2030[11] considers towns and rural areas important cogs in the planning of a better future for South Africans.

South African small town research has been limited in scope and scale.[12] It has for the most part been region specific and place specific, done by a limited number of researchers and with a glaring absence of contributions from parts of the country where a large number of small towns are located. The economic restructuring of rural South Africa in the light of changes in agriculture and closure of mines resulted in the majority of investigations being concerned with overviews and analyses of locally driven strategies for economic development. The restructuring of the South African economy in general, and the rural economy in particular, to a more post-productive landscape, intensified research on small towns and their hinterlands.[15]

Many small town investigations have been framed as single case studies, often with a focus on issues such as job creation through small, micro and medium enterprises (SMMEs), Public Works programmes and tourism development.[12] Local economic development (LED) provided a strategic vantage point from which to analyse small towns.[12,16] For instance, developmental LED policies and practices[12,16], infrastructure development[17], the development of small town tourism[18], and the impact of the changing fortunes of the mining sector received attention. Yet results were disappointing. Boom and bust cycles of mining resulted in many settlements being shadows of their former economic selves.[12]

Surprisingly, recent reviews of South African towns[12,16] did not pay much attention to quantitative research that revealed a strong orderliness in and predictability about the enterprise dynamics of these towns[2-4,9,10,19-24]. The latter studies have led to conclusions that challenge some long-held tenets about LED and job creation.[19,20] The former reviews and much of the research on which they are based have mostly been descriptive or narrative in nature. The understanding of what drives small towns in South Africa lacks the type of quantitative contribution that the MacArthur–Wilson Species Equilibrium Model has made to natural ecology.[18] A better understanding and control of the dynamics of small town growth and decline requires more holistic and quantitative and less descriptive/narrative research.

The aforementioned studies of the orderliness (regularities or proportionalities) in enterprise development and dynamics in towns moved in this direction. The regularities (proportionalities) have been detected as statistically significant correlations between entrepreneurial (e.g. total number of enterprises in towns or number of enterprises in certain business sectors in towns), economic (e.g. gross regional value added [GRVA], total regional personal income and regional employment) and demographic (e.g. town populations) characteristics in regional studies in South Africa. These regularities are expressions of an orderliness that prevails in enterprise development in South African towns. They indicate that in specific locations there are definitive limits to what can be attained and what not. The orderliness, therefore, confers quantitative predictive powers about enterprise development and dynamics of South African towns. Daniel Kahneman, a psychologist and economics Nobel laureate, lauded the predictive powers of simple algorithms, which according to him are often more accurate than the predictions of experts.[25] An economic systems model that links these regularities has been proposed for South Africa.[26] The present contribution enhances the understanding of the enterprise richness power law, adds predictive power, contributes to the information pool on which the model is based and could guide LED decision-making.

This discussion provides an affirmative answer to the first question: it is important to study the enterprise dynamics of South African towns.

Regional analysis

To answer the second question about the geographical spread of the power law requires comparison of different regions and information about the enterprise structures and dynamics of towns in those regions. National and regional information about enterprise structures is lacking in most countries. South Africa is probably the only country for which detailed enterprise structures and dynamics have been determined for different regions, i.e. towns in the Free State[9], the Karoo[6] and the GCBR[10].

The Free State

In the 1830s, European farmers disillusioned with the British authorities departed from the Cape Colony in the Great Trek.[27] Many settled in the sparsely inhabited area north of the Orange River on land traditionally occupied by African tribes. After the so-called Bloemfontein Convention in 1854 this area became the Republic of the Orange Free State, and later one of the adversaries in the South African War. Its rural areas were devastated during this war. The province is mostly a high plain centrally located in South Africa, and close on 130 000 km^2 in area. Livestock and grain farming formed the mainstay of the economy until gold was discovered in the mid-1940s.[28] A coal-to-oil plant and a new town, Sasolburg, followed in the 1950s. It has a rainfall gradient from the southwest to the northeast and in 2011 the population was 2.475 million. In 1910, when the Union of South Africa was formed, the Free State became the Orange Free State Province.[27] Currently known as the Free State, it is one of the nine provinces of the 'New South Africa'.

Most towns in the Free State were founded between the 1830s and 1910[29] and grew around newly built churches ('church towns') and some were located on important connection routes (e.g. Bethlehem and Kroonstad). Some on the eastern border, e.g. Ficksburg and Lady Brand, were founded for security reasons following the Basuto War of 1867. The discovery of gold in the mid-1940s and subsequent development of mines in the western Free State resulted in the founding of some new towns (e.g. Welkom, the second largest settlement in the province, and Allanridge) and the rapid growth of some existing towns (e.g. Odendaalsrus, Virginia and Hennenman). Sasolburg was specially founded in 1954 in the northern Free State to house workers of Sasol, a new coal-to-oil factory. Over time it became the third largest settlement in the Free State.[29]

A 2012/2013 study of the enterprise structures of Free State towns entailed 77 towns and villages (hereafter referred to as towns) and the city of Bloemfontein.[9] The ages of the towns and their geographical spread over the province differed widely. These data are used in this contribution and provide a strict test for the presence of a power law.

The Karoo

The Karoo is a vast arid region located in the centre of South Africa, comprising about 400 000 km^2, i.e. about 40% of South Africa's land surface.[30] It is a vast inland desert, with characteristic small, hardy, deciduous shrubs, low and variable rainfall (ranging from less than 100 mm per annum in the winter-rainfall west to 500 mm in the summer-rainfall east), occasional spectacular thunderstorms, and about 60 towns typically located 60–80 km from one another.

Karoo towns were slowly established from the late 1700s and throughout the 1800s; many were originally 'church towns' that served surrounding farm communities.[30] The Karoo economy was based on livestock, including cattle, goats and sheep. The region benefitted from the diamond and gold booms in the 1870s and 1880s, but because of drought, overgrazing, economic depression and the ravages of the South African War, a long-term economic decline began in the early 20th century. Recently, some Karoo towns have grown economically, while others have declined, showing a relatively uneven developmental profile.[13]

The Karoo is quite different from most of the Free State and, therefore, the enterprise structures of its towns would be useful in a comparison of regional enterprise richness power laws.

The Gouritz Cluster Biosphere Reserve

The GCBR is globally unique: it is the only area in the world where three recognised biodiversity hotspots – the Fynbos, Succulent Karoo and Maputaland-Pondoland-Albany – hotspots converge.[10] It also includes three recognised UNESCO World Heritage sites: the Swartberg Complex, Boosmansbos Nature Reserve and the western portion of the Baviaanskloof. It is located in the southern Cape of South Africa and covers a large area from the coast across the Langeberg Mountains to the Little Karoo, and even partly across the Swartberg Mountains. It includes 16 towns, which vary markedly in size.[10] Mossel Bay and Oudtshoorn are the two largest towns and Witsand, a small coastal town at the mouth of the Breede River, is the smallest. Several important routes cross the biosphere: the N2 and R62 from west to east, and the N9 and N12 (both national roads) from south to north. Route 62 is a well-known tourist route that passes through the Little Karoo.

The GCBR is also quite different from the Karoo and Free State and would also be useful in a comparison of regional enterprise richness power laws needed to answer the second question.

Time-based analysis of enterprise structures

To answer the third question about the durability of the power law, two sets of analyses of the same towns – but from different periods – were needed. The year 1946 was a special year in the history of the Free State. It was just after the end of World War II and its impacts, and just before gold mining in the Free State commenced with consequential economic development.[27] The donation of a telephone directory for 1946/1947 enabled an assessment of whether a power law relationship existed at that time between the enterprise richness and total enterprise numbers in Free State towns. This proved to be the case. Together with the analysis of Free State towns in 2012/2013[9], a comparison of enterprise richness power laws of the same region some seven decades apart was therefore possible, and which included a period of significant economic development in the Free State. This comparison would answer the third question.

Methods and results

Regional durability

Enterprise dynamics in Free State towns in 2012/2013

A first regional analysis involved data for 2012/2013 of 77 Free State towns (Table 1) as well as the provincial capital Bloemfontein (a metropolitan area). The data set formed part of an earlier study of the enterprise dynamics of Free State towns.[9]

Table 1: Total number of enterprises and enterprise types (enterprise richness) in Free State towns in 2012/2013

Town	Enterprises	Enterprise richness	Town	Enterprises	Enterprise richness	Town	Enterprises	Enterprise richness
Allanridge	18	14	**Hennenman**	120	56	**Rouxville**	29	21
Arlington	11	10	**Hertzogville**	26	20	**Sasolburg**	848	198
Bethlehem	993	224	**Hobhouse**	8	7	**Senekal**	139	72
Bethulie	38	31	**Hoopstad**	86	54	**Smithfield**	30	20
Bloemfontein	6617	429	**Jacobsdal**	40	29	**Soutpan**	7	6
Boshof	36	29	**Jagersfontein**	20	16	**Springfontein**	20	18
Bothaville	253	99	**Kestell**	37	27	**Steynsrus**	29	21
Botshabelo	203	73	**Koffiefontein**	39	30	**Thaba ‘Nchu**	129	52
Brandfort	87	45	**Koppies**	68	43	**Theunissen**	86	46
Bultfontein	108	52	**Kroonstad**	701	179	**Trompsburg**	42	29
Clarens	132	55	**Ladybrand**	258	104	**Tweeling**	14	14
Clocolan	75	46	**Lindley**	35	27	**Tweespruit**	20	16
Cornelia	6	6	**Luckhoff**	15	14	**Ventersburg**	47	27
Dealesville	11	11	**Marquard**	61	39	**Vierfontein**	21	14
Deneysville	72	41	**Memel**	32	27	**Viljoensdrif**	18	14
Dewetsdorp	29	23	**Odendaalsrus**	189	81	**Viljoenskroon**	143	68
Edenburg	26	20	**Oranjeville**	24	19	**Villiers**	56	37
Edenville	11	10	**Parys**	505	165	**Virginia**	313	108
Excelsior	21	17	**Paul Roux**	17	14	**Vrede**	140	75
Fauresmith	20	17	**Petrus Steyn**	41	27	**Vredefort**	41	26
Ficksburg	254	100	**Petrusburg**	33	27	**Warden**	41	31
Fouriesburg	51	32	**Philippolis**	28	20	**Welkom**	1837	272
Frankfort	171	77	**Phuthaditjhaba**	409	97	**Wepener**	37	29
Gariep Dam	22	13	**Reddersburg**	26	18	**Wesselsbron**	102	54
Harrismith	482	146	**Reitz**	133	68	**Winburg**	67	39
Heilbron	139	72	**Rosendal**	11	9	**Zastron**	80	43

The enterprises of each town were identified from the 2012/2013 Free State telephone[31] directory by previously described methods[2,3,19]. The type of each of the 17 184 enterprises recorded in the 78 settlements was identified from a list (at the time of writing of this document) of more than 670 enterprise types encountered in South African towns. The total number of enterprise types in each town represents its enterprise richness (Table 1).

The total enterprises to enterprise richness ratio of the Free State towns in 2012/2013 varied from 6:6 for Cornelia to 6617:429 for Bloemfontein (Table 1), thus covering a wide range of ratios. The enterprise richness in 2012/2013 of Free State towns exhibited a statistically significant ($p<0.01$) log-log (i.e. power law) relationship with total enterprise numbers (Figure 1). Just over 97% of the variance was explained by the power law which extended over a range from 6 to more than 6600 enterprises per settlement (Table 1). This result supports recent findings about the enterprise richness of South African towns.[4]

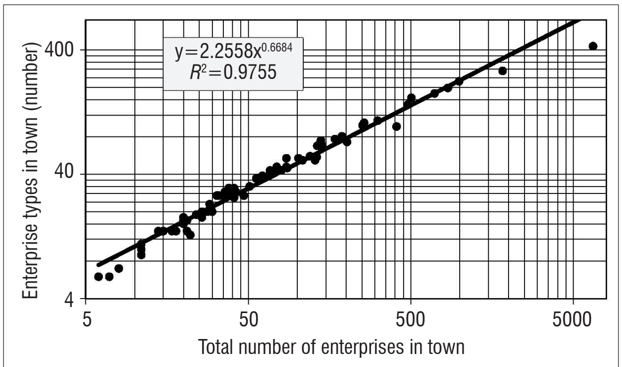


Figure 1: The relationship between the number of enterprise types (i.e. the enterprise richness) and the total number of enterprises in Free State towns and villages in 2012/2013.

Enterprise dynamics in Karoo towns

The Karoo is subject to a number of development and research initiatives, which include construction of the Square Kilometre Array (SKA) observatory, various solar power installations, power lines and potential shale gas production.[6] The dynamics of the enterprise richness of Karoo towns, i.e. whether they grow or regress, are therefore of interest. An earlier examination of the enterprise richness in 2014–2016 of 42 Karoo towns (Table 2) reported a statistically highly significant ($p<0.01$) power law relationship:

$$\text{Enterprise richness in town (no.)} = 1.4762x^{0.7655} \quad \text{Equation 1}$$

with $r=0.98$ and $n=42$, and where x is the total number of enterprises in the town.[12] This was the first indication that the enterprise richness of towns in the semi-arid to arid Karoo region of South Africa is a statistically significant ($p<0.01$) power law function of the total enterprises of the towns of the region.

The ratios of total enterprises to enterprise richness of the Karoo towns varied from 7:7 for Loxton to 353:104 for Beaufort West (Table 2). The power law relationship was, therefore, determined over a wide range of town sizes and was used in 'what-if' analyses of potential economic and entrepreneurial impacts of shale gas production on Karoo towns.[6]

Enterprise dynamics in 2013/2014 of towns of the GCBR

The GCBR is the only biosphere reserve in South Africa of which the enterprise structures of its towns have been analysed.[10] There are 16 towns of varying sizes in the GCBR (Table 3). Their total enterprises to enterprise richness ratios varied from 61:11 for Witsand to 1949:276 for Mossel Bay.

Enterprise richness in 2013/2014 of GCBR towns was also a statistically significant ($p<0.01$) power law function of the total enterprise numbers:

$$\text{Enterprise richness in town (no.)} = 1.2.684x^{0.7546} \quad \text{Equation 2}$$

with $r=0.93$ and $n=16$, and where x is the total number of enterprises in the town.

Almost 87% of the variance was explained by the power law which was estimated over a wide range of town sizes: 61 enterprises in Uniondale and Witsand to almost 1950 enterprises in Mossel Bay (Table 3 and Equation 2). The findings also support recent results about the enterprise richness of South African towns.[4]

The combined results of the three regional analyses demonstrate that power laws between enterprise richness and total enterprise numbers are not only characteristic of South African towns in general, or of groups of towns with similar enterprise structures, but also of regions in South Africa. This finding provided an answer to the second question.

Table 2: Total number of enterprises and enterprise types (enterprise richness) in 42 Karoo towns[6]

Town	Enterprises	Enterprise richness	Town	Enterprises	Enterprise richness	Town	Enterprises	Enterprise richness
Aberdeen	46	35	**Jansenville**	49	28	**Prieska**	108	56
Beaufort-West	353	104	**Kenhardt**	29	23	**Prince Albert**	83	42
Brandvlei	22	16	**Klipplaat**	13	11	**Richmond**	30	17
Britstown	27	17	**Laingsburg**	56	30	**Somerset East**	194	92
Calvinia	110	56	**Loeriesfontein**	29	22	**Steynsburg**	38	27
Carnarvon	58	37	**Loxton**	7	7	**Steytlerville**	34	23
Colesberg	144	58	**Middelburg**	161	80	**Strydenburg**	17	14
Cradock	294	109	**Murraysburg**	21	16	**Sutherland**	35	21
De Aar	223	89	**Nieu-Bethesda**	21	9	**Vanwyksvlei**	8	7
Fraserburg	33	22	**Nieuwoudtville**	30	19	**Venterstad**	18	14
Graaff-Reinet	347	130	**Noupoort**	30	23	**Victoria West**	74	43
Hanover	22	13	**Pearston**	20	12	**Vosburg**	9	7
Hofmeyr	17	12	**Petrusville**	17	14	**Williston**	26	21
Hopetown	70	49	**Philipstown**	15	13	**Willowmore**	53	27

Table 3: Total number of enterprises and enterprise types (enterprise richness) in Gouritz Cluster Biosphere Reserve towns in 2013/2014[10]

Town	Enterprises	Enterprise richness
Albertinia	114	55
Barrydale	82	30
Calitzdorp	72	33
De Rust	54	24
Great Brak River	170	80
Heidelberg	122	62
Ladismith	124	55
Montagu	269	94
Mossel Bay	1949	276
Oudtshoorn	1000	206
Prince Albert	155	59
Riversdal	283	114
Stilbaai	255	89
Swellendam	398	132
Uniondale	61	34
Witsand	61	11

The enduring nature of the enterprise richness of towns in South Africa

Two approaches were used to examine the enduring nature of the enterprise richness–total enterprises relationship: (1) using data obtained from a 1946/1947 telephone directory to determine if a power law also described the relationship at that time and then to compare the power laws of 1946/1947 and 2012/2013 and (2) using the 1946/1947 power law to predict the current enterprise richness of towns in the Free State, the Karoo and the GCBR.

For 1946/1947, 72 towns in the Free State, including the provincial capital Bloemfontein, were considered (Table 4). A list of all enterprises in each of the towns was prepared from a telephone directory for 1946/1947.[32] For provinces (including the Free State) other than the Cape Province, the directory contains a list of all trades (i.e. enterprises) for each town. The enterprise type of each of the 4177 enterprises listed in the 72 towns was identified from a list of more than 670 enterprise types encountered in South African towns. The number of enterprise types in each town represents its enterprise richness (Table 4).

The total enterprises to enterprise richness ratio of the Free State towns in 1946/1947 varied from 3:3 for Deneysville to 947:136 for Bloemfontein (Table 4). In the seven decades after 1946, the total number of enterprises in Free State towns had grown approximately fourfold, in part reflecting the stimulation of entrepreneurship by the development of the Free State goldmines and the coal-to-oil industry located in Sasolburg.[9]

Table 4: Total number of enterprises and enterprise types (enterprise richness) in Free State towns (towns and villages) in 1946/1947

Town	Enterprises	Enterprise richness	Town	Enterprises	Enterprise richness	Town	Enterprises	Enterprise richness
Bethlehem	225	73	**Hobhouse**	23	16	**Rosendal**	22	16
Bethulie	45	24	**Hoopstad**	26	16	**Rouxville**	35	22
Bloemfontein	947	136	**Jacobsdal**	13	10	**Senekal**	91	44
Boshof	46	26	**Jagersfontein**	34	20	**Smithfield**	37	22
Bothaville	65	28	**Kestell**	31	19	**Springfontein**	20	12
Brandfort	67	32	**Koffiefontein**	27	19	**Steynsrus**	42	22
Bultfontein	22	14	**Kopjes**	46	23	**Thaba 'Nchu**	48	27
Clarens	8	5	**Coalbrook**	10	5	**Theunissen**	46	23
Clocolan	64	38	**Kroonstad**	284	89	**Trompsburg**	45	27
Cornelia	12	11	**Ladybrand**	86	40	**Tweeling**	16	12
Dealesville	16	12	**Lindley**	50	30	**Tweespruit**	27	15
Deneysville	3	3	**Luckhoff**	16	10	**Ventersburg**	34	18
Dewetsdorp	37	22	**Marquard**	57	29	**Vierfontein**	7	5
Edenburg	31	19	**Memel**	18	13	**Viljoensdrif**	7	5
Edenville	20	14	**Odendaalsrus**	20	14	**Viljoenskroon**	48	20
Excelsior	11	9	**Oranjeville**	5	4	**Villiers**	36	23
Fauresmith	26	16	**Parys**	113	49	**Virginia**	8	5
Ficksburg	119	54	**Philippolis**	36	24	**Vrede**	77	39
Fouriesburg	33	17	**Paul Roux**	21	13	**Vredefort**	36	22
Frankfort	69	33	**Petrus Steyn**	27	17	**Warden**	38	21
Harrismith	129	54	**Petrusburg**	25	16	**Wepener**	57	30
Heilbron	70	39	**Witzieshoek**	8	4	**Wesselsbron**	27	18
Hennenman	34	22	**Reddersburg**	33	21	**Winburg**	72	34
Hertzogville	23	15	**Reitz**	90	37	**Zastron**	65	31

The 1946/1947 enterprise richness of Free State towns was a statistically significant ($p<0.01$) power law function of their total enterprise numbers. Just over 96% of the variance was explained by the power law which extended over a range from 3 to close on 950 enterprises per town (Figure 2). Seven decades ago, entrepreneurial space in Free State towns increased or decreased similarly to modern times as town sizes respectively increased or decreased.

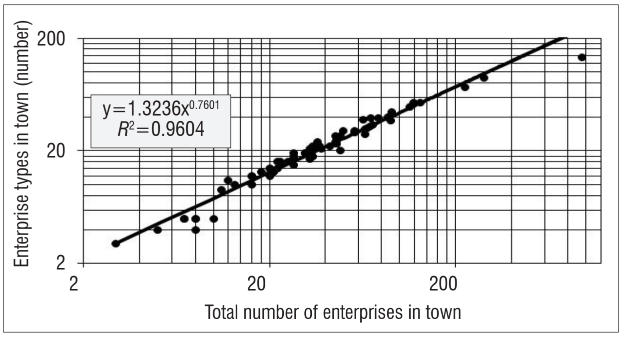


Figure 2: The relationship between the number of enterprise types (i.e. the enterprise richness) and the total number of enterprises in Free State towns and villages in 1946/1947.

As the size of towns (measured by their total enterprise numbers) increases, enterprise types appear that have not been present before. However, there are also more enterprises of enterprise types already present in the town. Toerien and Seaman[4] interpreted the power law in terms of how entrepreneurial space is created or destroyed as towns grow or regress. They distinguish between 'new entrepreneurs', who are those starting enterprises of enterprise types not yet present in towns, and 'existing entrepreneurs' who are those starting new enterprises of types already present in a town.

A graphic technique based on these considerations was developed to illustrate the roles of these two types of entrepreneurs in Karoo towns.[6] The same technique is used here to examine the enduring nature of the enterprise type–total enterprise power law. Using the power laws for 1946/1947 and 2012/2013, the relative roles of new and existing entrepreneurs were plotted over a wide range of town sizes (measured in number of enterprises) (Figure 3). There is a remarkable similarity in the definition of the entrepreneurial spaces in towns in the Free State seven decades apart. The way in which entrepreneurial space increased or decreased as towns respectively grew or regressed did not change much over seven decades.

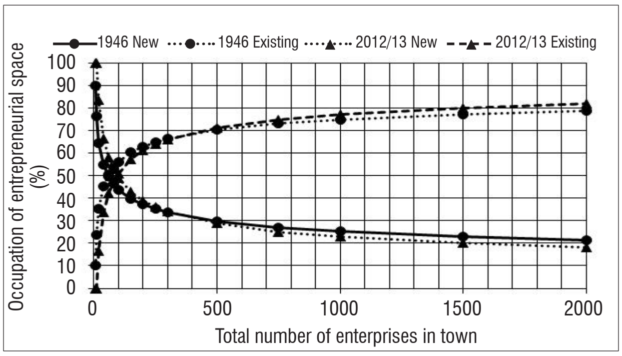


Figure 3: The entrepreneurial spaces for new and existing entrepreneurs in Free State settlements in 1946/1947 and 2012/2013.

New entrepreneurs are more important in small towns (with fewer than 60 to 100 enterprises) than are existing entrepreneurs (Figure 4). Small towns are expected to contribute to the growth of employment in South Africa. This expectation is clearly a huge challenge given that there are limitations in human capital and that the new entrepreneurs have to identify opportunities that are not obvious and cannot use 'me too' strategies. As towns grow, the importance of new entrepreneurs reduces: at approximately 60 to 100 enterprises the new to existing entrepreneurs ratio is 50:50 and thereafter the importance of existing entrepreneurs (and 'me too' strategies) increases, reaching about 80% at around 1800 enterprises. However, under all conditions, enterprise growth in Free State towns remains partly dependent on new entrepreneurs. For instance, this dependency is about 20% at 1800 or more enterprises per town. The graphic analysis is one illustration of the enduring nature of enterprise richness dynamics.

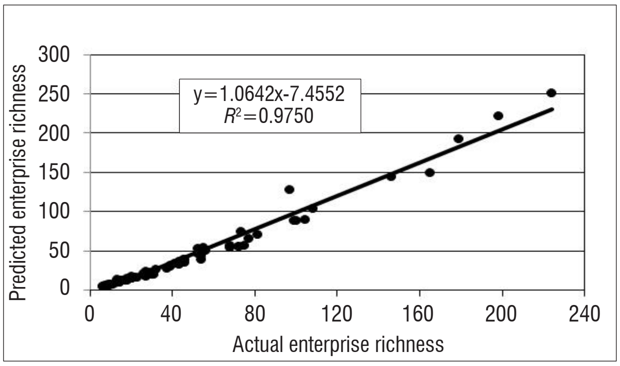


Figure 4: Enterprise richness in Free State settlements predicted by the 1946/47 power law versus actual enterprise richness measurements for 2012/2013.

The durability of the enterprise richness–total enterprises power law can also be tested by examining the ability of the 1946/1947 power law to predict the enterprise richness of Free State towns in 2012/2013 (Figure 4), in Karoo towns (Figure 5), and in GCBR towns in 2013/2014 (Figure 6). The comparisons used town size ranges limited to the maximum size of Free State towns in 1946/1947 (i.e. 947 enterprises in Bloemfontein). In each case, the actual number of enterprises of a town was used with the 1946/1947 power law to predict the expected enterprise richness. Best-fit regression lines of actual and predicted enterprise richness were calculated using Microsoft Excel.

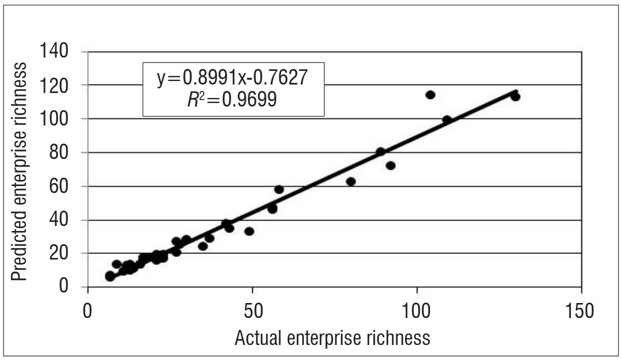


Figure 5: Enterprise richness in Karoo towns predicted by the 1946/47 power law versus actual enterprise richness measurements in recent times.

The expected enterprise richness was statistically significantly ($p<0.01$) predicted in every case (Figures 4, 5 and 6). Over seven decades, in different regions of the country and with vastly different socio-economic and political conditions in South Africa, the enterprise richness–total enterprises relationship has essentially stayed the same. This analysis answered the third question and provided a powerful demonstration of the universality of the enterprise richness power law in South Africa: entrepreneurial space develops or disappears in a regular way as towns grow or regress.

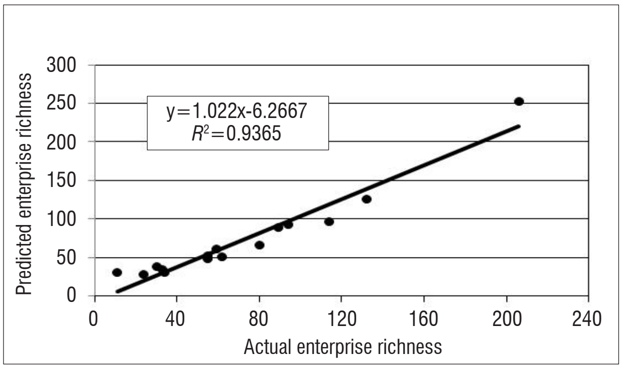


Figure 6: Enterprise richness in Gouritz Cluster Biosphere Reserve towns predicted by the 1946/1947 power law versus actual enterprise richness measurements for 2013/2014.

Discussion and conclusions

Empirical regularities in economics and finance often take the form of power laws.[33] Power laws were also detected in this study, and led to several important conclusions. Firstly, the presence of power laws serves as further proof of the presence of regularities in the enterprise dynamics of South African towns. The recorded power laws are very similar to each other. In fact, so similar that a power law extracted from seven-decade-old information could accurately predict the modern enterprise richness of regional South African towns (Figures 4, 5 and 6). The universality of enterprise richness power laws in South Africa indicates that entrepreneurial space in towns develops or disappears in a regular way as towns grow or regress.

The observed power laws form part of a set of a number of regularities observed in the enterprise dynamics of South African towns.[20,22-24] These regularities, not taken into account in recent reviews of South African towns[12,16], define the 'playing field' of enterprise development in South African towns, and set the boundaries of what is possible in terms of enterprise development and what is not. For instance, this study provides quantified insight into and predictive power about the role of new and existing entrepreneurs in the growth and decline of South African towns – concepts that are apparently alien to current wisdom about LED in South Africa. Development plans that are out of sync with these realities, are bound to fail. Research and planning of local economic and enterprise development in South Africa should move beyond descriptive or narrative approaches to more quantitative approaches.

Since Darwin's time, islands have provided laboratories for the study of evolution, including examining changes following colonisation.[34] Islands are ecological systems; and eventually the MacArthur–Wilson Species Equilibrium Model was developed as a fundamental basis of island biogeography.[18] It was shown that the species equilibria of islands are functions of island size and distance from mainlands.

Beinhocker[1] stressed the similarities between natural and economic systems. South African towns have been shown to be enterprise ecosystems[2] and to exhibit 'island effects' in the way enterprises develop in their domains[3]. This study has added to these ideas and provides quantified evidence that specific ratios of new and existing enterprise types are associated with specific sizes of towns (when expressed as total enterprises). It is an intriguing possibility that, analogous to the Species Equilibrium Model of natural ecology, it might be possible to develop an enterprise richness equilibrium model for South African towns. This possibility should receive further research attention.

The National Development Plan 2030[11] considers how two central South African problems, namely a lack of employment and poverty, could be addressed. The National Planning Commission considers towns and rural areas to be important cogs in such efforts.[11] The reshaping of South Africa's cities, towns and rural villages will be a complex and long-term project that requires major reforms, political will and differentiated planning responses in relation to varying town types.[11] An important aspect of discussions about the growth or regression of towns and the creation or disappearance of employment opportunities is enterprise structures and dynamics as towns grow or regress. This study supplies quantified information that should be applied.

Finally, it is necessary to consider why the enterprise richness power laws have apparently first been detected in South Africa. A possible reason is that in most countries there is limited, if any, information available on the enterprise richness of local economies. For instance, the US Census Bureau presents information on the number of firms (enterprises) on a state, county, city or town basis[35], but the type of each business is not recorded. In South Africa, the use of telephone directories to identify and enumerate enterprises in towns not only enabled their classification into 19 functional groups (e.g. enterprises involved in construction or trade services or tourism and hospitality services[19-23]) but also classification into more than 670 enterprise types.[4] The resulting unique database offers a unique opportunity to study and apply the enterprise richness relationships of South African towns.

Acknowledgements

Belinda Gordon is thanked for the donation of the 1946/1947 telephone directory. Marie Toerien and Estelle Zeelie provided technical assistance.

References

1. Beinhocker E. The origin of wealth: Evolution complexity, and the radical remaking of economics. Boston: Harvard Business School Press; 2006.

2. 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

3. 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

4. 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

5. Ball P. Critial mass: How one thing leads to another. Random House: London; 2005.

6. 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.

7. Spellerberg IF, Fedor PJ. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index. Glob Ecol Biogeogr. 2003;12:177–179. http://dx.doi.org/10.1046/j.1466-822X.2003.00015.x

8. May RM. Foreword. In: Losos JB, Ricklets ER, editors. The theory of island biogeography revisited. New Jersey: Princeton University Press; 2010. p. vii–ix.

9. Toerien DF. The enterprise architecture of Free State towns. Technical report 2014. http://dx.doi.org/10.13140/2.1.1902.7521

10. Toerien DF. Enterprise profiles and dynamics in a South African biosphere reserve. Technical report 2016. http://dx.doi.org/10.13140/RG.2.1.1434.4083

11. South African National Planning Commission (NPC). Our future – make it work. National Development Plan 2030. Pretoria: NPC; 2011.

12. Hoogendoorn G, Visser G. South Africa's small towns: A review on recent research. Local Econ. 2016;31:95–108. http://dx.doi.org/10.1177/0269094215618865

13. Nel E, Hill T, Taylor B, 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. http://dx.doi.org/10.1007/s12132-011-9131-z

14. Harrison P, Todes A. Spatial transformations in a 'loosening state': South Africa in a comparative perspective. Geoforum. 2015;61:148–162. http://dx.doi.org/10.1016/j.geoforum.2015.03.003

15. Hoogendoorn G, Nel E. Exploring small town development dynamics in rural South Africa's post-productivist landscapes. In: Donaldson R, Marais L, editors. Small town geographies in Africa: Experiences from South Africa and elsewhere. New York: Nova Science Publishers; 2012. p. 21–34.

16. Rogerson CM, Nel E. The contested trajectory of applied local economic development in South Africa. Local Econ. 2016;31:109–123. http://dx.doi.org/10.1177/0269094215623729
17. Rogerson CM. Reframing place-based economic development in South Africa: The example of local economic development. Bull Geogr Socio-Econ Ser. 2014;24:203–218. http://dx.doi.org/10.2478/bog-2014-0023
18. Schoener TW. The MacArthur–Wilson Equilibrium Model: A chronicle of what it said and how it was tested. In: Losos JB, Ricklets ER, editors. The theory of island biogeography revisited. New Jersey: Princeton University Press; 2010. p. 52–87.
19. Toerien DF, Seaman MT. Ecology, water and enterprise development in selected rural South African towns. Water SA. 2011;37(1):47–56. http://dx.doi.org/10.4314/wsa.v37i1.64106
20. 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
21. Toerien DF, Seaman MT. Regional order in the enterprise structures of selected Eastern Cape Karoo towns. S Afr Geogr J. 2012;94(2):1–15. http://dx.doi.org/10.1080/03736245.2012.742782
22. 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
23. Toerien DF. 'n Eeu van orde in sakeondernemings in dorpe van die Oos-Kaapse Karoo [A century of order in businesses in towns of the Eastern Cape Karoo]. LitNet Akad. 2014;11(1):330–371. Afrikaans.
24. Toerien DF. Economic value addition, employment, and enterprise profiles of local authorities in the Free State, South Africa. Cogent Soc Sci. 2015;1(1), Art. #1054610, 21 pages. http://dx.doi.org/10.1080/23311886.2015.1054610
25. Kahneman D. Thinking, fast and slow. Kindle ed. London: Penguin; 2013.
26. 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
27. Giliomee H, Mbenga B. New history of South Africa. Cape Town: Tafelberg; 2007.
28. Free State Provincial Government. Free State Growth and Development Strategy: Free State vision 2030 – the future we want. Bloemfontein: Free State Provincial Government; 2013.
29. Erasmus BPJ. Op pad in Suid-Afrika [On the road in South Africa]. Johannesburg: Jonathan Ball; 1995. Afrikaans.
30. Atkinson D. Is South Africa's Great Karoo region becoming a tourism destination? J Arid Environ. 2016;127:199–210. http://dx.doi.org/10.1016/j.jaridenv.2015.12.006
31. Free State 2012/2013 telephone directory. Johannesburg: Trudon; 2012.
32. Cape Times Ltd. South African directory 1946-47. Cape Town: Cape Times Ltd.; 1946.
33. Gabaix X. Power laws in economics and finance. Annu Rev Econ. 2009;1:255–293. http://dx.doi.org/10.1146/annurev.economics.050708.142940
34. Losos JB, Ricklefs RE. Foreword. In: Losos JB, Ricklefs RE, editors. The theory of island biogeography revisited. Princeton, NJ: Princeton University Press; 2010. p. xi–xiv.
35. United States Census Bureau. Quick facts: United States [database on the Internet]. No date [cited 2016 Aug 30]. Available from: https://www.census.gov/quickfacts/table/PST045215/00

Emission factors of domestic coal-burning braziers

AUTHORS:
Tafadzwa Makonese[1]
Daniel M. Masekameni[2]
Harold J. Annegarn[1,3]
Patricia B.C. Forbes[4]

AFFILIATIONS:
[1]Sustainable Energy Technology and Research Centre, University of Johannesburg, Johannesburg, South Africa
[2]Department of Environmental Health, University of Johannesburg, Johannesburg, South Africa
[3]Energy Institute, Cape Peninsula University of Technology, Cape Town, South Africa
[4]Department of Chemistry, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Tafadzwa Makonese

EMAIL:
taffywandi@gmail.com

DATES:
Received: 23 June 2016
Revised: 01 Oct. 2016
Accepted: 14 Oct. 2016

KEYWORDS:
emission rates; domestic coal combustion; air quality; ventilation rates; ignition method

HOW TO CITE:
Makonese T, Masekameni DM, Annegarn HJ, Forbes PBC. Emission factors of domestic coal-burning braziers. S Afr J Sci. 2017;113(3/4), Art. #2016-0187, 11 pages. http://dx.doi.org/10.17159/sajs.2017/20160187

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
University of Johannesburg; Global Alliance for Clean Cookstoves; South African National Energy Development Institute (SANEDI); GTZ

We present experimental results of emission factors from a suite of domestic coal-burning braziers (lab fabricated and field collected) that span the possible range of real-world uses in the Highveld region of South Africa. The conventional bottom-lit updraft (BLUD) method and the top-lit updraft (TLUD) method were evaluated using coal particle sizes between 20 mm and 40 mm. Emission factors of CO_2, CO and NO_x were in the range of 98–102 g/MJ, 4.1–6.4 g/MJ and 75–195 mg/MJ, respectively. Particulate matter ($PM_{2.5}$ and PM_{10}) emissions were in the range 1.3–3.3 g/MJ for the BLUD method and 0.2–0.7 g/MJ for the TLUD method, for both field and lab-designed stoves. When employing the TLUD method, emission factors of $PM_{2.5}$/PM_{10} reduced by up to 80% compared with those when using the BLUD method. Results showed the influence of ventilation rates on emission factors, which reduced by ~50% from low to high ventilation rates. For energy-specific emission rates, the combined (3-h) PM_{10} emission rates were in the range of 0.0028–0.0120 g/s, while the combined average CO emission rates were in the range of 0.20–0.26 g/s, with CO_2 emission rates in the range of 0.54–0.64 g/s. The reported emission factors from coal braziers provide the first comprehensive, systematic set of emission factors for this source category, and fill a major gap in previous efforts to conduct dispersion modelling of South African Highveld air quality.

Significance:

- The study provides the first comprehensive, systematic set of emission factors from coal braziers.
- The study fills a major gap in previous efforts to conduct dispersion modelling of South African Highveld air quality.
- Results have implications for stove design and lay the groundwork for improvements in the design of existing coal braziers.
- Results have implications for understanding the potential health impacts of condensed matter emissions from coal braziers.

Introduction

Exposure to fine particulate matter (PM) from solid fuel combustion is associated with morbidity[1-3] and mortality[4-5], especially in the developing world[6]. About three billion people worldwide are exposed daily to harmful emissions from the combustion of solid fuels. Combustion of these fuels releases products of incomplete combustion such as carbon monoxide, PM and volatile organic compounds.[7] The WHO Global Health Observatory has reported that household air pollution caused the premature deaths of ~4.3 million people globally in 2012, while a further 3.7 million premature deaths were attributable to ambient air pollution.[8] Household air pollution is associated with many health effects such as acute and chronic respiratory disorders, and pulmonary and systemic diseases.[7]

Emission inventories for PM are critical in establishing how sources affect health and climate and, therefore, need to be continuously improved.[6] To date, emission factors from the literature are still adopted in the development of emission inventories. However, there is currently a lack of sufficient and reliable data, especially for emission factors, which leads to uncertainties and bias in many emission inventories due to influences of a variety of parameters.[6,9] For example, combustion technology and operational practice of appliances have a major influence on the physicochemical properties of the emitted particles.[6,10-13] Reported emission factors from domestic burning vary as a result of differences in (1) fuel properties (e.g. moisture and volatile matter content); (2) stove design; (3) fire ignition methods (top-lit versus bottom-lit); (4) fire management and ventilation (e.g. air supply amount and fuel-air mixing condition); and (5) experimental methods (e.g. laboratory chamber, simulated stove/open burning and field measurement).[6,9,14-17]

In South Africa, particularly on the interior Highveld plateau, combustion of coal in open braziers is among the largest sources of PM and black carbon emissions. Coal is still used as a primary cooking and heating fuel for the majority of the population.[18] There are two ignition methods for lighting a coal fire, namely the traditional bottom-lit updraft (BLUD) method and the reduced smoke top-lit updraft (TLUD) method, locally known as the *Basa njengo Magogo* (BnM).

The rapid electrification of households after 1990, now reaching over 95% of urban dwellings, did not result in the anticipated switch away from combustion of solid fuels for heating and cooking.[19,20] As an interim method to reduce population exposure to domestic smoke emissions, the South African Department of Energy embarked on a public awareness campaign to encourage the dissemination and uptake of the BnM method as a no-cost method of reducing smoke emissions.[21,22] The TLUD approach is a simple intervention in the manner in which residential fires are lit, and involves placing kindling at the top rather than at the bottom of the fuel load in a brazier or stove. In contrast, the traditional BLUD method involves placing the kindling and igniting it at the bottom of the brazier or stove, with the bulk of the fuel placed on top of the burning kindling.[23-25] The TLUD method is estimated to result in a 70–90% reduction in ambient particulate emissions and a 20% reduction in coal consumption at no additional cost to households.[26,27] For several years (from 2009 to 2014), the implementation of the BnM rollout became a national priority energy intervention programme.

However, these reports had limitations that reduce confidence in the reported emission factors – the studies did not control or systematically evaluate braziers regarding fuel quality, ventilation rates or fuel moisture content, leading to substantial uncertainties in derived emission inventories.[27,28] Determination of such influences is necessary for an improved understanding of emission characteristics and more accurate emission factors.[6,29] In 1997, the Atomic Energy Corporation of South Africa characterised one standard coal sample and 10 low-smoke test fuels for physical, chemical and pollutant emission factors.[27] The authors used only a single stove in the experiments but noted that there are many different stoves and brazier designs in use in South Africa, and that brazier designs would influence emission factors. As such, it is likely that specific fuels would have altered emission profiles when combusted in different braziers.

We carried out a comprehensive and systematic study of gaseous and PM emissions from the combustion of commonly used D-grade coal in typical artisanal coal-burning braziers. The information provided in this article will be useful for improving emission inventories for CO, CO_2, NO_x and PM from domestic coal combustion.

Materials and methodology

Experimental stoves

Brazier stoves are made by hand out of metal drums with perforations of varying sizes punched around the sides, with a wire grate installed across the middle to support the fuel bed. There is no standard brazier design, as the devices vary widely regarding the number, size and distribution of the side holes (which affects ventilation rates), as well as the presence or absence of a grate and its position in the drum. Braziers are used widely in informal settlements and less affluent suburbs for heating and roadside cooking while burning wood, coal or even trash.

Figure 1: Photographs of experimental field procured braziers with (a) high ventilation, (b) medium ventilation and (c) low ventilation; and experimental lab-designed braziers with (d) high ventilation, (e) medium ventilation and (f) low ventilation.

Three braziers (*imbaula*), part of a set of 11 procured from user communities (hereafter field stoves) (Figure 1a–c) and three stoves designed and built at the SeTAR laboratory at the University of Johannesburg (hereafter lab stoves) (Figure 1d–f) were evaluated for emissions performance. Tests were conducted under laboratory-controlled conditions at the SeTAR Centre. The lab stoves had uniformly sized and spaced holes, designed so that the total ventilation hole area approximated the corresponding field stoves – high, medium and low ventilation cases, respectively. By the nature of the *imbaula* device, the irregular shape and packing density of coal pieces within a particular combustion sequence does not allow for precise measurement or replication of air flow through the holes into the combustion zone. While the lab stove makes an attempt to replicate the 'as found' artisanal field stoves by reproducing the same ventilation hole area in a regularly spaced array of uniform size holes, complete congruency could not be established between the field and lab stoves.

Stove ventilation rates

Stove ventilation rates, as a function of air hole density, can affect the overall performance of a fuel/stove combination. Ventilation rates differ from one brazier model to another depending on the hole size, hole configuration (or pattern) and hole density. Ventilation rates need to be measured and specified to be able to evaluate and compare the performance of two or more braziers.[11] The total hole areas, indicative of the ventilation rates of the braziers, for the field and lab stoves are given in Table 1.

Lab stoves were fabricated by drilling 20-mm-diameter air holes on new, empty 20-L metal paint drums (H=360 mm, D=295 mm); different hole densities were employed for the high, medium and low ventilation cases, and holes were distributed above and below the grate. The grate was set at 120 mm from the base – a height typical of the prototype braziers collected from the informal settlement.[30] The hole sizes in the field stoves varied from 10 mm to 40 mm; a uniform hole diameter of 20 mm was used for the lab stoves.

Fire ignition method

The BnM fire ignition method is a local derivative of the TLUD procedure. The sequence for setting a TLUD fire is to place the major portion of the coal load on the fuel grate, and then place kindling (paper and wood chips) on top, ignite the kindling, and finally put a few coal nuggets on top of the burning kindling a few minutes after ignition. As such, about 2.0 kg of coal was added onto the fuel grate, followed by 0.4 kg of kindling (0.04 kg newspaper and 0.36 kg wood chips). After ignition, ~1.0 kg of coal is added on top of the burning kindling.[11] For the BLUD, the sequence is reversed: ~1.0 kg of coal is placed onto the grate, followed by 0.4 kg of kindling. After ignition, ~2.0 kg of coal is added to the burning kindling. The BLUD method results in copious smoke emissions, as the fire is oxygen-starved during the initial stages of combustion.[11]

Fuel characterisation

The coal fuel used in the experiments was purchased from a local coal merchant and compared with coal sourced directly from a colliery in the Witbank Emalahleni coalfield. The initial field survey revealed that local merchants obtained a D-grade type coal fuel from Slater Coal Mine (Dundee, KwaZulu-Natal, South Africa). In light of the above, batches of D-grade coal, sufficient to conduct a range of comparative tests, were purchased from the coalfield. For each fuel batch, the fuel was first mixed for homogeneity and then three 2-kg samples were taken to an independent laboratory for analyses. Calorific value, proximate (moisture, ash, volatile organic compounds, fixed carbon) and ultimate (%C, %H, %S, %O and mineral elements) analyses were performed (Table 2).

The coal was crushed and sieved. The fraction passing through a 40-mm upper sieve and retained by a 20-mm lower sieve was used in all experiments to avoid variability from coal particle size – a variable that was not investigated in this study. The fuel was then stored in moisture-free containers. Immediately before commencing a series of combustion tests, a sample of the fuel was re-analysed for moisture content using an electric oven. The quantities of fuel used in the experiments are detailed in the section on the fire ignition method.

Table 1: Ventilation rates of the experimental stoves

	Field stoves			Lab stoves		
	High ventilation	**Medium ventilation**	**Low ventilation**	**High ventilation**	**Medium ventilation**	**Low ventilation**
Air hole area above the grate (cm^2)	159	166	91	126	101	63
Air hole area below the grate (cm^2)	248	189	63	138	50	38
Total air hole area (cm^2)	407	355	154	264	151	101

Table 2: Fuel analysis for the D-grade coal used in the experiments

Parameter (air-dried basis)	Standard method	Slater colliery D-grade coal
Moisture content (%)	ISO 5925	3.50
Volatiles (%)	ISO 562	20.30
Ash (%)	ISO 1171	24.20
Fixed carbon (%)	By difference	52.00
Calorific value (MJ/kg)	ISO 1928	23.40
Calorific value (Kcal/kg)	ISO 1928	5590
Total sulfur (%)	ASTM D4239	0.63
Carbon (%)	ASTM D5373	62.60
Hydrogen (%)	ASTM D5373	2.72
Nitrogen (%)	ASTM D5373	1.43
Oxygen (%)	By difference	4.96

Sampling procedure

Gas measurement techniques

In this study, the hood method[31] was used for monitoring emissions from domestic coal-burning braziers. Because the experimental braziers did not have a flue, the stoves were placed under a smoke collection hood attached to the SeTAR dilution system responsible for the ducting and dilution of the exhaust gas stream.[11] When using an extraction fan, high extraction rates tend to influence the combustion characteristics of a stove.[32] In light of this, an extractor fan was not used for drawing air through the hood and duct. An extractor fan may be useful in preventing flue gases escaping from the bottom of the hood.[32] However, our system relies on the carbon mass balance method, which only requires a representative sample of the exhaust gases to quantify the emissions and not the total capture of all combustion products. An advantage of the hood method is that it can be employed for the simultaneous determination of thermal and emission parameters in a standard and systematic manner.[33] Figure 2 shows the experimental set-up.

The measurement instruments (gas and particle analysers) used in this study cannot cope with the dense smoke emitted from braziers during the initial phases of combustion. This dense smoke saturates the analytical instruments, which results in underreporting of emission factors as a result of the concentrations exceeding the measurement range of the instruments, or causes the machines to clog with condensed organic deposits, resulting in erroneous readings. The SeTAR dilution system was designed with these consequences in mind. The dilution system is connected to the hood and the sampling nozzle is placed in the stack gas.

The inlet of the sampling nozzle is not equipped with a $PM_{2.5}$ fractionating cyclone, as the majority of the particles are condensation or Aitken particles below 1 μm in diameter. This sampling nozzle is followed by a mixing tunnel in which the hot exhaust is diluted with aerosol-free filtered air. The diluted exhaust then passes through a 0.9-m long ageing chamber (mixing chamber), and then through sampling ports for gaseous and particle emissions monitoring. The dilution level can be pre-set or changed by continuously varying the volume of dry air, depending on need. By recording diluted CO_2 and undiluted CO_2, the dilution ratio can be calculated for each 10-s interval. As such, there is no need for flow monitoring or calibration of flow meters. The volume of dry CO_2-free air supplied through the jet can be controlled by using a rotameter with a needle-valve flow controller.

The sampling configuration for the undiluted gas channel comprised Teflon tubing, a 4-μm pore diameter filter pack, and a flue gas analyser (Testo® 350XL/454) with cells for CO_2, CO, NO_x and O_2.[34] For the diluted channel, the sampling configuration included the dilution system, a Teflon tube, and a flow splitter to duct gas streams to a DustTrak DRX 8533 aerosol monitor and a second Testo® analyser.

Particle measurement techniques

The particle mass concentration is determined in the unfiltered, diluted air stream. The sampling configuration for PM included the SeTAR dilution system and the particulate monitor (DustTrak DRX 8533 aerosol monitor), connected by Teflon tubing (Figure 3). The DustTrak aerosol monitor is an optical particle counter that simultaneously measures size-segregated mass fraction concentrations (PM_1, $PM_{2.5}$, PM_4, PM_{10} and total particle mass) in real time over a 0.001–150 mg/m^3 concentration range. The instrument is calibrated by the manufacturer, Raeco, annually. It is noted that the mass of the sub-micron fractions reported by optical particle counters may be overestimated in comparison with gravimetric methods.[35] PM emission factors reported in this paper are based on the mass determined from the optical particle counter, and no post-measurement correction factors have been applied to compensate for possible overestimations. While relevant for purposes of dispersion modelling, any discrepancies in the mass will not invalidate the comparisons made between ignition methods and ventilation rates.

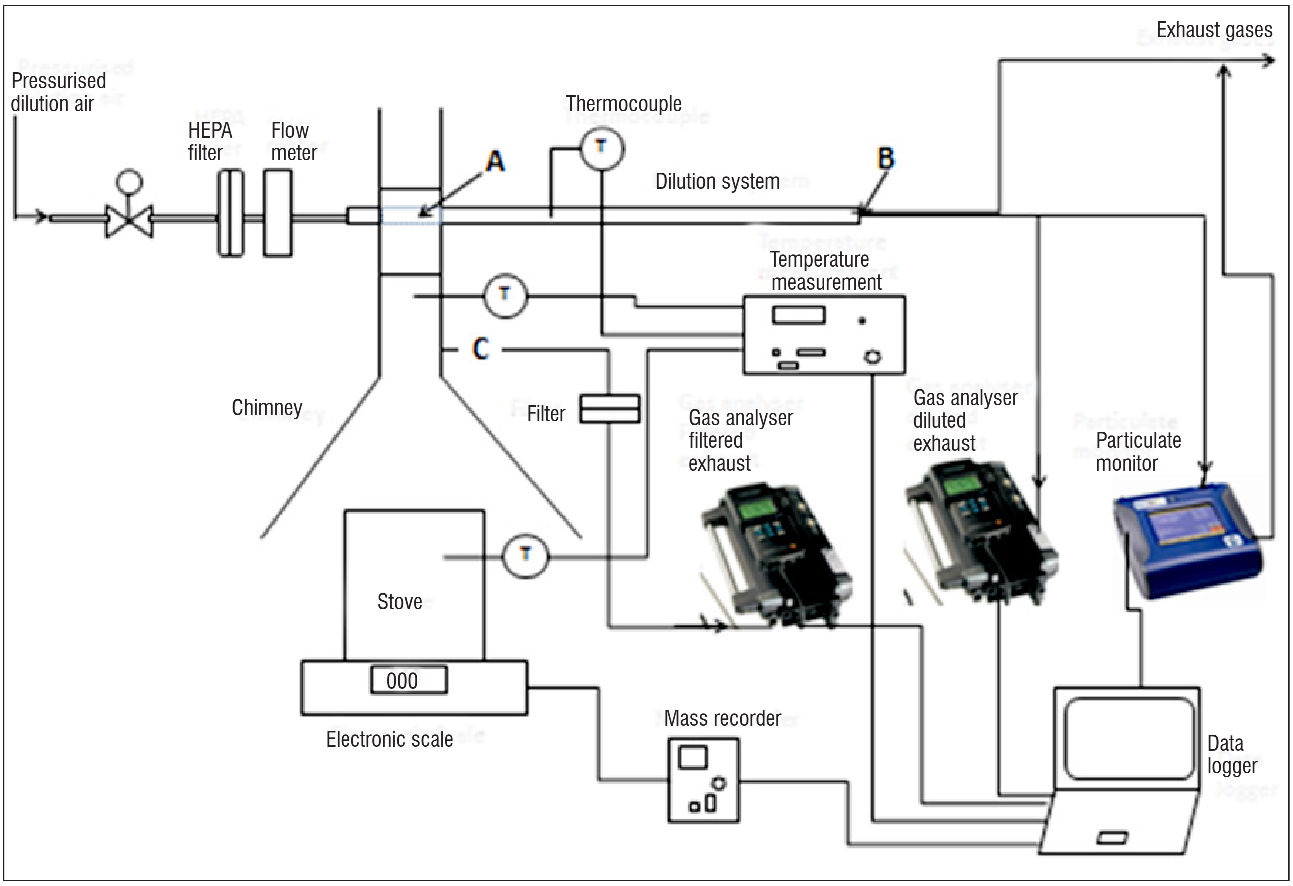


Figure 2: Schematic illustration of the SeTAR stove testing rig, showing the mixing point (A) where the exhaust is mixed with filter compressed air for dilution; and the sampling point (B) where the diluted exhaust is channelled to the monitoring instruments.

Calculations

Gas emission factors

The measurement of emissions using the hood method was carried out on three field and three lab stoves that span the range of braziers used in the informal settlements of Gauteng Province. As in Bhattacharya et al.[32], concentrations of CO, CO_2 and NO_x were monitored, enabling calculation and estimation of emission factors for each pollutant.

Emission factors presented in this study are calculated as in Bhattacharya et al.[32] with slight modifications (methane and non-methane hydrocarbons are not reported herein), for energy-specific emission factors in units (g/MJ) of energy in the fuel. The lower heating value is used for calculations of thermal efficiency numbers and the determination of energy-specific emission factors. For each test, CO, CO_2 and NO_x concentrations (ppmv) were recorded every 10 s for the duration of the burn sequence or the test experiment.

Using the carbon mass balance method, CO and CO_2 emission factors were calculated using:

$$C_{moles} = (C_{mass} – C_{ash})/ M_{carbon} \qquad \text{Equation 1}$$

where C_{moles} is the total moles of carbon burnt, C_{mass} is the mass of carbon in the fuel, C_{ash} is the mass of carbon in the ash and M_{carbon} is the molecular mass of carbon.[32]

C_{moles} in the flue gas can be determined by:

$$C_{moles} = \eta CO + \eta CO_2 \qquad \text{Equation 2}$$

where ηCO refers to the moles of carbon in CO in the flue exhaust, and ηCO_2 to the moles of carbon in CO_2 in the flue exhaust.

$$\eta CO_2 = (C_{moles})\, CO_2\, (CO + CO_2)^{-1} \qquad \text{Equation 3}$$

$$\eta CO = (C_{moles})\, CO\, (CO + CO_2)^{-1} \qquad \text{Equation 4}$$

Mass-specific emission factors (EF) for CO_2 and CO in units (g/kg) are given by:

$$CO_2EF = \eta CO_2 \text{ x } M_{CO2} \text{ (mass of fuel consumed)}^{-1} \qquad \text{Equation 5}$$

$$COEF = \eta CO \text{ x } M_{CO} \text{ (mass of fuel consumed)}^{-1} \qquad \text{Equation 6}$$

where M_{CO2} and M_{CO} are the molecular masses of CO_2 and CO, respectively.

Energy-specific emission factors for CO_2 and CO expressed in units (g/MJ) are given by:

$$CO_2EF = \eta CO_2 \text{ x } M_{CO2} \text{ (net heat gained)}^{-1} \qquad \text{Equation 7}$$

$$COEF = \eta CO \text{ x } M_{CO} \text{ (net heat gained)}^{-1} \qquad \text{Equation 8}$$

where the net heat gained refers to the heat retained by a cooking vessel during the water heating experiments.

Excess air

Excess air was determined using the SeTAR data calculation sheet, which employs a chemically balanced approach (which is detailed on the SeTAR website). The determination of the total air demand (λ) is given by:

$$\lambda = 1 + \frac{O_{2\,meas} - O_{2\,oxid}}{O_{2\,det} - (O_{2\,meas} - O_{2\,oxid})} \qquad \text{Equation 9}$$

where $O_{2\,meas}$ is the concentration of O_2 measured in the exhaust; $O_{2\,oxid}$ is the amount of O_2 required to oxidise completely unburned gases or products of incomplete combustion; and $O_{2\,det}$ is the total O_2 in all detected gases.[18]

Energy-specific particulate emission factors

Gaseous emission (in ppmv) can be converted to other units including energy-specific emission factors referenced to the energy content in the fuel consumed, namely g/MJ and mg/MJ. There is a need to first determine the net heat gained from the fuel. Net heat gained is[23]:

> *The heat retained by a cooking vessel during a burn sequence and is expressed in units of megajoules (MJ). It includes the heating of the pot and its contents plus the heat of evaporation of water, but excludes other heat flows through the pot, specifically radiative and convective losses from the pot sides and top.*

The mass of detected $PM_{2.5}$ or PM_{10} is first multiplied by the dilution factor and then multiplied by λ to obtain the actual mass emitted before dilution. According to Language et al.[35], this approach is based on the foreknowledge that any missing fuel has been turned into combustion products of some type. This method can track and correctly determine the performance of the stove in real time while burning fuels in an inhomogeneous manner, as is often the case with biomass and coal fuels. The standard reporting metrics for the particle mass concentration include the mass of PM emitted per net megajoule of energy delivered into the pot, or mass of PM emitted per net megajoule of energy delivered from the fire.[36] For example, the mass of PM_{10} emitted during a burn sequence is determined and divided by the net heat gained, yielding emissions per net megajoule (g/MJ) or (mg/MJ). For example:

$$PM10_{EF} = \frac{PM10[\text{g}]}{H_{NET}[\text{MJ}]} \qquad \text{Equation 10}$$

Emission rates

Emission rates (g/s) for PM and gaseous emissions were determined, averaged per hour for each hour of a maximum 3-h burn sequence, and also averaged over the first 3 h of the burn sequences.

Statistical analyses

An F-test was used to test whether different stove test results for the ignition method and ventilation rates had the same variance. Because the F-test is a relatively robust statistical tool, we used high alpha levels (0.05) and balanced layouts. The F-test result was then used to determine the best *t*-test analysis to use on the data (i.e. a two-sample *t*-test, assuming equal variances or a two-sample *t*-test, assuming unequal variances).

A two-tailed Student *t*-test (at the 95% confidence level) was used for evaluating the thermal and emissions data between fire ignition methods and between ventilation rates. For this study, a statistically significant value was taken as a *p*-value less than 5% ($p<0.05$).

Quality control

For each brazier model, a series of preliminary burn sequences was carried out to standardise procedures and to minimise the variability from differences in user/operator behaviour. After that, five definitive tests were conducted for each brazier model. After every test run, the gas probes and Teflon tube channels were cleaned, and the pumps and machines checked and zeroed.

Continuous gas and particle monitoring instruments are routinely sent for calibration at intervals prescribed by the manufacturers, or at least once annually, and need to be periodically verified with laboratory standards. Zero and span calibration were performed on all analysers before and after every test run to account for small variations in the dilution ratio. For example, the DustTrak DRX was zeroed with filtered air before each test run.

Before conducting test experiments, the sampling dilution system components were disassembled, cleaned, air dried, assembled, and

Table 3: Particulate matter (PM) and gas emission factors for the lab-designed coal braziers using D-grade coal, for the top-lit updraft (TLUD) and bottom-lit updraft (BLUD) ignition methods at low, medium and high ventilation rates ($n=5$)

		BLUD method		TLUD method		Statistical analysis		
Pollutant	**Ventilation rate**	**Emission factor**	**s.d.**	**Emission factor**	**s.d.**	**Difference between BLUD and TLUD (%)**	***t*-statistic**	***p*-value**
$PM_{2.5}$(g/MJ)	High	1.3	0.1	0.3	0.02	-80%	22	0.002*
	Medium	1.9	0.1	0.4	0.03	-76%	16	0.004*
	Low	2.5	0.2	0.6	0.06	-76%	13	0.006*
PM_{10}(g/MJ)	High	1.3	0.1	0.3	0.02	-80%	22	0.002*
	Medium	1.9	0.1	0.4	0.03	-76%	16	0.004*
	Low	2.5	0.2	0.6	0.06	-76%	13	0.006*
CO(g/MJ)	High	4.5	0.1	4.6	0.2	2%	-0.61	0.61
	Medium	6.2	0.3	5.0	0.4	-20%	3.3	0.08
	Low	6.4	0.5	5.7	0.2	-11%	2.1	0.17
CO_2(g/MJ)	High	98	6	99	2	-1%	0.36	0.75
	Medium	98	6	96	2	-2%	0.46	0.69
	Low	97	5	102	3	5%	-1.0	0.41
NO_x(mg/MJ)	High	149	16	126	14	-16%	1.6	0.26
	Medium	118	11	99	6	-16%	2.2	0.16
	Low	87	5	75	5	-13%	2.5	0.13

*$p<0.05$

'tested for leaks to prevent contamination from the surrounding air'[37] to avoid contamination of the emissions with organic and metal compounds from previous burn sequences. High power compressed air and water were used to remove large particles from the sampling channels. The exhaust collection trains, involving stainless steel ducts, Teflon tubes, and sampling nozzles, were cleaned with soap and water and air dried with filtered compressed air.

The response time of monitoring instruments can be affected by the time the exhaust takes to travel from the inlet probe to the instrument and through the sensing volume within the instrument. The time lags were measured before and after a test. The system was first allowed to run on air for a few minutes to sample ambient air for purposes of obtaining background readings. Then an ignition torch was lit near the inlet of the sampling probe. After the ignition torch had been extinguished, the system sampled ambient air for a few more minutes to allow the instrument signals to return to background levels, after which the ignition torch was lit again. These time delays were noted and then corrected for in the spreadsheet when calculating pollutant emission factors for different fires.

Results and discussion

Gaseous emission factors

General observations

When the kindling was lit for the BLUD method, the coal immediately began to give off sulfurous odours and dense whitish/yellowish smoke – a consequence of devolatilised organic matter that had not reached combustion temperature or had insufficient oxygen to oxidise fully. A few minutes elapsed before the coal could be considered ignited (taken as parts of the lowest lumps of coal glowing visibly red). However, the thick white smoke continued for up to 30 min as volatiles were evolved from the gradually heated coal above the rising combustion front. There was considerably less smoke for the TLUD ignition method, with high flames above the coal bed resulting from ignition of the evolved devolatilised organic matter passing through the descending combustion front, with homogenous phase combustion continuing in the rising gas mixture above the coal bed.

Gas and PM emission factors for lab stoves using D-grade coal

Gas and PM emission factors for the laboratory-designed stoves, using D-grade coal, for TLUD and BLUD methods are presented in Table 3, together with a statistical comparison of the differences. The emission factors are based on integration over the combustion sequence, from ignition until 3 h had elapsed, or fuel burnout, whichever was the soonest.

For any given ventilation rate, and for either TLUD and BLUD ignition, $PM_{2.5}$ and PM_{10} emission factors were similar. Particles <2.5 μm in aerodynamic diameter constitute more than 90% of the mass of particulate emissions <10 μm in aerodynamic diameter (Table 3). There were significant differences ($p<0.05$) in particle emissions between the BLUD and TLUD methods, for all three ventilation rates. The TLUD approach reduced $PM_{10}/PM_{2.5}$ emissions by 76–80% compared to the 'business as usual' BLUD method. This result is similar to assertions made by Le Roux[26] who found a reduction in particulate emissions of between 78% and 92% when using the TLUD ignition method.

Comparison of ignition methods (Table 3) shows that there were no significant differences ($p>0.05$) in CO, CO_2 and NO_x emissions at any given stove ventilation rate, implying that the ignition method did not affect the combustion characteristics of the pollutants in the devices tested. There were no significant differences between NO_x emissions when varying the ignition method, although there was an average 16% decrease in the emissions for the TLUD method. Results for NO_x are

Table 4: Particulate matter (PM) and gas emission factors for the field stoves using D-grade coal, for the top-lit updraft (TLUD) and bottom-lit updraft (BLUD) ignition methods at low, medium and high ventilation rates ($n=5$)

		BLUD method		TLUD method		Statistical analysis		
Pollutant	Ventilation rate	Emission factors	s.d.	Emission factors	s.d.	Difference between BLUD and TLUD (%)	*t*-statistic	*p*-value
$PM_{2.5}$(g/MJ)	High	1.3	0.1	0.2	0.02	-81%	15	0.000*
	Medium	2.9	0.3	0.6	0.05	-78%	15	0.000*
	Low	3.3	0.2	0.7	0.06	-80%	23	0.000*
PM_{10}(g/MJ)	High	1.3	0.1	0.2	0.02	-81%	15	0.000*
	Medium	2.9	0.2	0.6	0.05	-78%	15	0.000*
	Low	3.3	0.2	0.7	0.06	-80%	23	0.000*
CO(g/MJ)	High	4.1	0.3	4.0	0.2	-3%	1.0	0.57
	Medium	4.2	0.4	4.1	0.4	-4%	1.0	0.57
	Low	4.6	0.3	5.5	0.2	20%	-4.0	0.02*
CO_2(g/MJ)	High	102	4	100	6	-2%	1.0	0.62
	Medium	102	5	99	3	-3%	1.0	0.49
	Low	98	6	101	5	2%	0.0	0.65
NO_x(mg/MJ)	High	195	8	168	10	-14%	4.0	0.02
	Medium	188	10	163	4	-13%	4.0	0.02
	Low	187	11	161	9	-14%	3.0	0.04

*$p<0.05$

as expected, as the maximum temperatures reached in the braziers for either ignition method were below the threshold temperature for oxidation of atmospheric nitrogen. Fuel-derived NO_x was released in proportion to concentration in the fuel and was invariant to the ignition method or ventilation rate.

Gas and PM emission factors from field stoves using D-grade coal

Gas and PM emission factors for the field designed stoves, using D-grade coal, for TLUD and BLUD ignition methods are presented in Table 4.

There were significant differences ($p<0.05$) in PM_{10} emission factors between the fire ignition methods. Emissions of $PM_{2.5}$ and PM_{10} from the TLUD ignition method were fivefold lower (80% reduction) than those from the BLUD method, across all ventilation rates. These results confirm results presented in Table 3 for the lab-designed stoves.

For CO, there was a significant difference in CO emissions between the fire ignition methods at low ventilation rates. At this ventilation rate, there was a 20% increase in CO when using the TLUD method. Lack of oxygen resulted in smouldering combustion conditions that favoured the emission of products of incomplete combustion, including CO and polycyclic aromatic hydrocarbons.[11] NO_x emissions show that a significant difference exists between ignition methods across all ventilation rates. Further investigations are required to investigate this anomaly.

Influence of stove ventilation rates

The effect of ventilation rates on the emissions of PM, CO, CO_2 and NO_x were investigated. Three stoves with different ventilation rates were used for each ignition method, with moisture content and other parameters held constant. The results of these experiments for the lab-designed braziers are presented in Table 5.

There were significant differences ($p<0.05$) in PM_{10} emissions for both ignition methods at low, medium and high ventilation rates, implying that ventilation rate affected the combustion characteristics of PM_{10} in all the experimental braziers. For a given device, switching from low to high ventilation rates reduced $PM_{2.5}$ and PM_{10} emission factors by 50%. This advantage is offset by a firepower excessive for convenient cooking (about 10 kW) especially during the ignition and pyrolisation combustion phases as well as by increased fuel consumption (by a factor of ~1.3 compared to low ventilation rates).

Table 5: Emission factors (mg/MJ) as a function of ventilation rates for the top-lit updraft (TLUD) and bottom-lit updraft (BLUD) ignition methods ($n=5$)

Ignition method	Ventilation rate	PM_{10}	CO	CO_2	NO_x
BLUD	High	1.3 ± 0.1*	4.5 ± 0.1*	98 ± 6	149 ± 16*
	Medium	1.9 ± 0.1	6.2 ± 0.3	98 ± 6	118 ± 11
	Low	2.5 ± 0.2	6.4 ± 0.5	97 ± 5	87 ± 5
TLUD	High	0.3 ± 0.02*	4.6 ± 0.2*	99 ± 2	126 ± 4*
	Medium	0.4 ± 0.03	5.0 ± 0.4	96 ± 2	99 ± 6
	Low	0.6 ± 0.06	5.7 ± 0.2	102 ± 3	75 ± 5

**significant difference between high and low ventilation rates*

However, this improvement from increased ventilation is not as significant as the decrease from the top-lit ignition. A well-ventilated brazier, when used in conjunction with the TLUD ignition method, has the potential to have the maximum reduction of PM emissions from coal fires in open braziers.

There was a significant difference ($p<0.05$) in CO emissions, but not CO_2 emissions, between the high and low ventilation rates. This finding shows that ventilation rate influences CO emissions, but may not influence the combustion characteristics of CO_2. There was also a significant difference in NO_x emissions between high and low ventilation rates for BLUD and TLUD fires.

Emission rates

The emission rates of CO, CO_2 and PM_{10} for selected braziers with three different ventilation rates, using D-grade coal, were calculated from measurements using the TLUD and BLUD ignition methods. Knowing the burn rate of the fuel (mass loss) and concentrations of the exhaust components, we determined the exhaust emission rates during the operation of the braziers, per hour of operation (Table 6) and for the full 3-h combustion sequence (Table 7). The emission rates are expressed in units (g/s), as required by dispersion model input specifications. Averages of five tests are given together with the standard deviation. Higher particulate emission rates occurred in the first hour of combustion (ignition and pyrolysis phases) than in the second and third hours, which can be associated with condensed semi-volatile compounds that are driven off from the kindling and during initial heating of the coal. Thereafter, the particulate emission rates decreased considerably, reaching lowest levels during the last hour of the 3-h combustion sequence. In constructing realistic emission patterns from domestic combustion for dispersion modelling, the correct diurnal time evolution of emissions becomes an important factor in generating accurate dispersion predictions. Such hourly emission factors can be adapted to the lifestyle patterns of the modelled communities when conducting health risk assessments based on dispersion modelling and exposure assessments.

Table 7 presents a summary of 3-h (full burn sequence) emission rates for lab-designed and field-obtained braziers for each ventilation rate using the BLUD and the TLUD fire ignition methods and the combined average emission rate over the three ventilation rates. There were no differences in the combined average CO_2 emission rates for the two ignition methods and between the lab and field stoves. The combined average CO_2 emission rates were 0.54–0.64 g/s. There were differences in the combined average CO emission rates between the stoves and the ignition methods. The combined average CO emission rates were in the range of 0.20–0.26 g/s. The combined average PM_{10} emission rates were in the range of 0.0028–0.012 g/s (Table 7).

The data presented in Table 6 and Table 7 are useful for short-term modelling of pollutants in which 1-h average emission rates are required inputs for standard dispersion models. Source data input requirements for air dispersion modelling include specific mass emission rates, physical stack measurements (e.g. diameter, stack height, exit velocity and temperature of the exhaust gases) and dry and wet deposition settling parameters for PM. For domestic sources that are too numerous to model individually, emissions are treated as area sources, therefore requiring a knowledge of the source density [number of homes/emission points per km^2] and the emission rate per source (g/s). The source density can be derived from remote-sensing images and GIS methods, or from census data. However, up to now, there has been no systematic study to characterise emissions from a range of domestic coal stoves and braziers, including variability of stove design, fuel quality, and stage of the burn or user behaviour. Emission values presented in Table 6 and Table 7 may be used to construct better area source emission estimates than have been available up to now, in the absence of reliable emission factors of individual coal combustion devices, and the time variations of these emissions.

Limitations of the research study

The primary purpose of this study was to investigate in a systematic manner some of the factors influencing *imbaula* emissions – namely ventilation rates and ignition methods. For this purpose, the number of replicates used in the study is deemed adequate to reach sound conclusions. As an adjunct, emission factors have been derived. The selection of field stoves was based on covering the range encountered in one township. This range included a wide variety of drum sizes, diameters and distributions of holes, and state of disrepair. A comprehensive survey would be required to determine the properties of the larger range of *imbaulas* used across the Highveld region.

Table 6: Hourly average emission rates (g/s) during burn sequence (for dispersion modelling input) for lab and field stoves for the top-lit updraft (TLUD) and bottom-lit updraft (BLUD) ignition methods at low, medium and high ventilation rates (*n*=5)

Ignition	Stove	Ventilation rate	Hour	Energy (MJ/h)	s.d. (MJ)	CO_2 (g/s)	s.d. (g/s)	CO (g/s)	s.d. (g/s)	PM_{10} (g/s)	s.d. (g/s)
BLUD	Lab	Low	1	6.39	0.51	1.11	0.09	0.031	0.003	4.1E-02	5.1E-03
			2	2.27	0.18	0.36	0.03	0.010	0.001	1.5E-04	3.8E-05
			3	1.41	0.07	0.22	0.01	0.011	0.001	4.5E-07	2.6E-08
		Medium	1	6.60	0.44	1.11	0.07	0.030	0.002	6.3E-02	6.1E-03
			2	2.27	0.17	0.38	0.03	0.020	0.002	1.4E-05	2.9E-06
			3	2.01	0.08	0.34	0.01	0.006	0.000	8.7E-07	3.6E-08
		High	1	5.30	0.80	0.84	0.13	0.046	0.007	2.0E-02	2.4E-03
			2	2.11	0.14	0.33	0.02	0.024	0.001	3.5E-04	5.6E-05
			3	1.32	0.07	0.19	0.01	0.023	0.001	1.6E-05	8.8E-07
	Field	Low	1	6.93	0.84	1.21	0.15	0.063	0.008	4.9E-02	5.9E-03
			2	2.27	0.11	0.38	0.02	0.010	0.001	5.5E-04	5.0E-05
			3	1.41	0.09	0.22	0.01	0.006	0.000	2.6E-05	1.4E-06
		Medium	1	7.00	0.70	1.34	0.13	0.043	0.007	3.4E-02	5.6E-03
			2	2.43	0.15	0.34	0.02	0.012	0.001	1.4E-05	1.1E-06
			3	1.95	0.16	0.24	0.02	0.012	0.001	7.1E-05	7.0E-06
		High	1	7.27	1.01	1.28	0.18	0.021	0.005	2.4E-02	4.1E-03
			2	2.55	0.13	0.44	0.02	0.027	0.002	2.0E-05	1.1E-06
			3	1.93	0.13	0.33	0.02	0.029	0.002	9.2E-05	1.6E-05
TLUD	Lab	Low	1	8.60	0.79	1.46	0.15	0.055	0.007	1.3E-02	1.1E-03
			2	1.86	0.12	0.29	0.02	0.014	0.001	1.5E-05	1.6E-06
			3	1.18	0.06	0.18	0.01	0.013	0.001	9.9E-06	6.9E-07
		Medium	1	3.82	0.92	0.67	0.16	0.036	0.008	7.9E-03	1.4E-03
			2	3.60	0.17	0.62	0.03	0.021	0.001	3.4E-03	2.2E-04
			3	3.48	0.20	0.60	0.03	0.027	0.002	2.0E-04	2.2E-05
		High	1	6.90	1.49	0.62	0.24	0.009	0.002	2.3E-03	1.0E-03
			2	2.70	0.20	0.42	0.03	0.028	0.002	7.0E-07	3.8E-08
			3	1.42	0.09	0.20	0.01	0.030	0.002	9.9E-07	6.6E-08
	Field	Low	1	6.60	1.18	1.13	0.20	0.030	0.009	1.4E-02	2.0E-03
			2	1.93	0.13	0.32	0.02	0.014	0.001	3.1E-05	3.6E-06
			3	1.20	0.07	0.19	0.01	0.015	0.001	9.2E-05	1.3E-05
		Medium	1	5.62	0.73	0.92	0.12	0.030	0.003	8.6E-03	8.4E-04
			2	2.51	0.16	0.39	0.03	0.025	0.002	4.1E-04	5.9E-05
			3	1.48	0.13	0.21	0.02	0.026	0.002	3.4E-05	3.0E-06
		High	1	4.35	0.66	0.76	0.12	0.023	0.003	2.1E-03	3.4E-04
			2	2.30	0.17	0.38	0.03	0.021	0.001	5.3E-05	8.1E-06
			3	1.37	0.09	0.20	0.01	0.023	0.002	6.1E-07	3.6E-08

Table 7: Average emission factors over a 3-h combustion sequence for lab and field stoves for the top-lit updraft (TLUD) and bottom-lit updraft (BLUD) ignition methods at low, medium and high ventilation rates (n=5)

Ignition	Stove	Ventilation rate	Energy (MJ) per 3-h	s.d. (MJ)	CO_2 (g/s)	s.d. (g/s)	CO (g/s)	s.d. (g/s)	PM_{10} (g/s)	s.d. (g/s)
BLUD	Lab	Low	3.36	0.25	0.56	0.04	0.017	0.0015	1.4E-02	1.7E-03
		Medium	3.63	0.23	0.61	0.04	0.019	0.0012	2.1E-02	2.0E-03
		High	2.91	0.34	0.45	0.05	0.031	0.0032	6.8E-03	8.2E-04
		Mean	**3.30**	**0.27**	**0.54**	**0.05**	**0.022**	**0.0020**	**1.4E-02**	**1.5E-03**
	Field	Low	3.54	0.35	0.60	0.06	0.026	0.0029	1.7E-02	2.0E-03
		Medium	3.79	0.34	0.64	0.06	0.022	0.0029	1.1E-02	1.9E-03
		High	3.92	0.42	0.69	0.07	0.026	0.0029	7.9E-03	1.4E-03
		Mean	**3.75**	**0.37**	**0.64**	**0.06**	**0.025**	**0.0029**	**1.2E-02**	**1.7E-03**
TLUD	Lab	Low	3.88	0.32	0.64	0.06	0.027	0.0027	4.2E-03	3.6E-04
		Medium	3.63	0.43	0.63	0.07	0.028	0.0037	3.8E-03	5.5E-04
		High	3.67	0.59	0.41	0.09	0.022	0.0020	7.8E-04	3.4E-04
		Mean	**3.73**	**0.45**	**0.56**	**0.08**	**0.026**	**0.0028**	**2.9E-03**	**4.2E-04**
	Field	Low	3.24	0.46	0.55	0.08	0.020	0.0037	4.7E-03	6.6E-04
		Medium	3.20	0.34	0.51	0.06	0.027	0.0021	3.0E-03	3.0E-04
		High	2.67	0.31	0.45	0.05	0.023	0.0019	7.2E-04	1.2E-04
		Mean	**3.04**	**0.37**	**0.50**	**0.06**	**0.023**	**0.0026**	**2.8E-03**	**3.6E-04**

Another limitation is that emissions from *imbaulas* should not be used as a surrogate for emissions from cast iron coal stoves that have entirely different ventilation structures, combustion characteristics and probably user behaviours. We used only a single grab sample of coal. Emissions rates may also vary according to the coal quality and/or from which mine or geological seam the coal was derived, and according to the type and quantity of kindling used to ignite the coal. However, determining the effects of coal quality and kindling on emissions was outside the scope of this study. We kept the kindling uniform in terms of quality and quantity throughout the experiments.

For experimental purposes, controlled methods of fire setting were adopted. No attempt was made to survey the range of user behaviours, or to model some of the extreme behaviours encountered in the field regarding setting the fuel load and ignition method. For example, roadside mealie cookers fill the *imbaula* to the brim (5 kg coal) before ignition, resulting in a prolonged period of intense smoke emission. As with most combustion stoves, user behaviour is one of the largest factors in variability in stove efficiency and emissions performance and is a topic worthy of separate investigation.

Conclusions and implications

This study is the most comprehensive systematic study to date of emission factors from domestic coal braziers in South Africa. The insights gained on the controlling variables go a long way to understanding the divergent results from prior studies that did not take into account or control for all these variables. We report emission factors regarding emissions per net MJ of energy in the fuel, for use in rating and comparison of stoves. The experimentally determined emission rates, in units (g/s), directly applicable in standard dispersion models, represent a considerable improvement in previously reported emission factors for coal stoves. These results have implications for stove designs, laying the groundwork for improvements in the design of existing coal braziers and development of novel low-emission combustion technologies, and understanding potential health impacts of condensed matter emissions.

Future improvements could explore additional factors, including the monitoring of organic carbon emissions that have the ability to influence the accuracy of the estimations of CO and CO_2 emission factors. Additional factors also include varying the size of ventilation holes and the partitioning of ventilation openings below the grate (primary air) and above the grate or above the fuel load (secondary air), and the mean particle size of the fuel. We kept the fuel median size and size distribution constant at 20–40 mm throughout the course of this study.

There is a need to investigate further the effect of hole density pattern on ventilation rates by clustering primary air holes below the fuel grate and a limited number of secondary air holes above the level of the packed fuel bed. Results presented herein show that the performance of braziers could be improved by optimising the air supply through the packed fuel bed and to the space above the fuel bed (secondary air). An increase in the ventilation rate allows for more stable combustion and higher combustion temperatures, which in turn could lead to improved heat transfer. Increased ventilation could be achieved by using forced draft as in some wood pellet and biomass burning gasifiers. However, this alternative is not viable for brazier stoves as it may require design alterations to existing braziers, which is not an option for artisanally produced stoves. The target market for these stoves remains the low-income (energy poor) households situated in the Highveld region of South Africa.

In conclusion, PM and trace gas emissions from coal braziers could be optimised by modifying the ignition procedure, the design of the braziers regarding hole distribution and diameter, and the position of the grate in the brazier. Condensed matter particulate (smoke) emissions can be reduced by allowing for thorough mixing of the volatile gases from the fuel bed and the air supply, and by allowing a long residence time in the high-temperature zone. Thus, a brazier with high ventilation rates has the potential to reduce emissions of CO and PM. Future studies on the effect of ventilation rates on fixed-bed coal combustion systems should explore parameters such as the size of ventilation holes, the distribution pattern of the holes, and fuel size on particulate and trace gas emissions, as well as the impact of user behaviour on emission rates.

Acknowledgements

This work was supported financially by the University of Johannesburg through a URC/Faculty of Science grant to the SeTAR Centre, and in part from a grant from the Global Alliance for Clean Cookstoves (GACC) to the SeTAR Centre as a Regional Stove Testing and Development Centre. Part of the equipment and early stages of this work were funded by a grant from the South African National Energy Development Institute (SANEDI) for the project: ENV-0607-128 'Emissions testing from the BNM (Basa Njengo Magogo) reduced smoke fire lighting method'. We thank the German GTZ who provided substantial funding for the development of the SeTAR Laboratory through a legacy grant from the Basic Energy and Climate Change Adaptation Programme (BECCAP). T.M. thanks the Faculty of Engineering and the Built Environment for a Postdoctoral Research Fellowship grant. We thank Thokozile Sithole for assistance with the combustion experiments and Crispin Pemberton-Pigott for part analysis of the combustion data.

Authors' contributions

T.M. performed all the experiments, analysed the data, carried out quality control and assurance of the data and stove testing rig, wrote the first draft of the manuscript, and edited the manuscript according to the reviewers comments and suggestions by the editorial team. D.M.M. assisted with the experiments, including collecting stove data from the field, analysed field and experimental data and edited the manuscript. H.J.A. supervised the project, provided the finances, analysed the data in terms of QC/QA, and edited the manuscript for content, language and grammar. P.B.C.F. supervised the project, analysed the experimental data, performed QC/QA on the data, and edited the manuscript for content, language and grammar.

References

1. Titcombe ME, Simcik M. Personal and indoor exposure to $PM_{2.5}$ and polycyclic aromatic hydrocarbons in the southern highlands of Tanzania. Environ Monit Assess. 2011;180(1–4):461–476. http://dx.doi.org/10.1007/s10661-010-1799-3
2. Naeher LP, Brauer M, Lipsett M, Zelikoff JT, Simpson CD, Koenig JQ, et al. Woodsmoke health effects: A review. Inhal Toxicol. 2007;19:67–106. http://dx.doi.org/10.1080/08958370600985875
3. Englert N. Fine particles and human health: A review of epidemiological studies. Toxicol Lett. 2004;149(1–3):235–242. http://dx.doi.org/10.1016/j.toxlet.2003.12.035
4. Barnes DF. Clean household and indoor air pollution. Washington DC: The World Bank; 2006.
5. Bruce N, Perez-Padilla R, Albalak R. Indoor air pollution in developing countries: A major environmental and public health challenge. Bull WHO. 2000;78:1078–1092.
6. Shen GF, Yang YF, Wang W, Tao S, Zhu C, Min Y, et al. Emission factors of particulate matter and elemental carbon for crop residues and coals burned in typical household stoves in China. Environ Sci Technol. 2010;44(18):7157–7162. http://dx.doi.org/10.1021/es101313y
7. Gordon SB, Bruce NG, Grigg J, Hibberd PL, Kurmi OP, Lam KH, et al*.* Respiratory risks from household air pollution in low and middle income countries. Lancet Resp Med. 2014;2(10):823–860. http://dx.doi.org/10.1016/S2213-2600(14)70168-7
8. World Health Organization (WHO). The global health observatory data repository: Population using solid fuels (estimates) [database on the Internet]. c2014 [cited 2014 Sep 16]. Available from: http://apps.who.int/gho/data/node.main.135?lang=en,2014
9. Shen G, Xue M, Wei S, Chen Y, Zhao Q, Li B, et al. Influence of fuel moisture, charge size, feeding rate and ventilation conditions on the emissions of PM, OC, EC, parent PAHs, and their derivatives from residential wood combustion. J Environ Sci. 2013;25(9):1808–1816. http://dx.doi.org/10.1016/S1001-0742(12)60258-7
10. Uski O. Toxicological effects of fine particles from small-scale biomass combustion [PhD thesis]. Joensuu: University of Eastern Finland; 2014.
11. Makonese T, Forbes P, Mudau L, Annegarn HJ. Monitoring of polycyclic aromatic hydrocarbon (PAH) emissions from real-world uses of domestic coal braziers. In: Proceedings of the Domestic Use of Energy (DUE) Conference; 2014 March 31 – April 02; Cape Town, South Africa. Cape Town: IEEE Conference Publications; 2014. p. 1–8. http://dx.doi.org/10.1109/DUE.2014.6827759
12. Tissari J, Lyyränen J, Hytönen K, Sippulaa O, Tapperb U, Frey A, et al*.* Fine particle and gaseous emissions from normal and smouldering wood combustion in a conventional masonry heater. Atmos Environ. 2008;42:7862–7873. http://dx.doi.org/10.1016/j.atmosenv.2008.07.019
13. Bølling AK, Totlandsdal AI, Sallsten G, Braun A, Westerholm R, Bergvall C, et al*.* Wood smoke particles from different combustion phases induce similar pro-inflammatory effects in a co-culture of monocyte and pneumocyte cell lines. Part Fibre Toxicol. 2012;9:45. http://dx.doi.org/10.1186/1743-8977-9-45
14. Chen YJ, Roden CA, Bond TC. Characterizing biofuel combustion with patterns of real-time emission data (PaRTED). Environ Sci Technol. 2012;46(11):6110–6117. http://dx.doi.org/10.1021/es3003348
15. Jetter J, Zhao Y, Smith KR, Khan B, Yelverton T, DeCarlo P, et al. Pollutant emissions and energy efficiency under controlled conditions for household biomass cookstoves and implications for metrics useful in setting international test standards. Environ Sci Technol. 2012;46(19):10827–10834. http://dx.doi.org/10.1021/es301693f
16. Johnson M, Edwards R, Berrueta V, Masera O. New approaches to performance testing of improved cookstoves. Environ Sci Technol. 2010;44:368–374. http://dx.doi.org/10.1021/es9013294
17. Jetter J, Kariher P. Solid-fuel household cook stoves: Characterisation of performance and emissions. Biomass Bioenerg. 2009;33:294–305. http://dx.doi.org/10.1016/j.biombioe.2008.05.014
18. Makonese T, Annegarn HJ, Masekameni D, Forbes P. Influence of fuel-bed temperatures on CO and condensed matter emissions from packed-bed residential coal combustion. In: Proceedings of the Domestic Use of Energy (DUE) Conference; 2015 March 31 – April 02; Cape Town, South Africa. Cape Town: IEEE Conference Publications; 2015. p. 63–69. http://dx.doi.org/10.1109/DUE.2015.7102964
19. Madubansi M, Shackleton CM. Changing energy profiles and consumption patterns following electrification in five rural villages, South Africa. Energ Policy. 2006;34(18):4081–4092. http://dx.doi.org/10.1016/j.enpol.2005.10.011
20. Davis M. Rural household energy consumption: The effects of access to electricity – Evidence from South Africa. Energ Policy. 1998;26(3):207–217. http://dx.doi.org/10.1016/S0301-4215(97)00100-6
21. Standish B, Leiman A, Boting A, Van Zyl H. Reducing the health care costs of urban air pollution: The South African experience. J Environ Manage. 2007;84:27–37. http://dx.doi.org/10.1016/j.jenvman.2006.05.010
22. Makonese T. Protocols for thermal and emissions performance testing of domestic fuels and stoves [MSc thesis]. Johannesburg: University of Johannesburg; 2011.
23. Makonese T, Masekameni MD, Forbes P, Annegarn HJ. Influence of fire-ignition methods and stove ventilation rates on gaseous and particle emissions from residential coal braziers. J Energy South Africa. 2015;26(4):16–28.
24. Wagner NJ, Schoonraad P, Swanepoel P, Van Niekerk A, Scholtz C, Kornelius G, et al. Results of domestic smoke reduction programmes at eMbalenhle (Mpumalanga) and Zamdela (Free State). In: Proceedings of the National Association for Clean Air (NACA) Annual Conference; 2005 September 29–30; Cape Town, South Africa. Vereeniging: NACA; 2005. p. 1–16. http://dx.doi.org/10.13140/2.1.2485.4725
25. South African Council for Scientific and Industrial Research (CSIR). Laboratory controlled quantitative information about reduction in air pollution using the "*Basa njengo Magogo*" methodology and applicability to low-smoke fuels (revised). Report no. ENV-D 2005-004. Durban: Division of Water, Environment & Forestry Technology, CSIR; 2005.
26. Le Roux LJ, Zunckel M, McCormick S. Reduction in air pollution using the 'Basa njengo Magogo' method and the applicability to low-smoke fuels. J Energy South Afr. 2009;20(3):3–10.
27. Van Niekerk WCA, Britton MTS, Laurens JB. Laboratory technical tests – Determination of emission factors. Pretoria: Department of Minerals and Energy; 1997.

28. Scorgie Y, Kneen MA, Annegarn HJ, Burger LW. Air pollution in the Vaal Triangle – Quantifying source contribution and identifying cost-effective solutions. Clean Air J. 2003;13(2):5–18.

29. Wei W, Zhang W, Hu D, Ou L, Tong Y, Shen G, et al. Emissions of carbon monoxide and carbon dioxide from uncompressed and pelletized biomass fuel burning in typical household stoves in China. Atmos Environ. 2012;56:136–142. http://dx.doi.org/10.1016/j.atmosenv.2012.03.060

30. Kimemia DK, Annegarn HJ, Robinson J, Pemberton-Pigott C, Molapo V. Optimising the imbaula stove. Paper presented at: The Domestic Energy Use (DUE) Conference; 2011 April 04; Cape Town, South Africa.

31. Ahuja DR, Veena J, Smith KR, Venkataraman C. Thermal performance and emission characteristics of unvented biomass-burning cook-stoves: A proposed standard method for evaluation. Biomass. 1987;12:247–270. http://dx.doi.org/10.1016/0144-4565(87)90039-4

32. Bhattacharya SC, Albina DO, Salam PA. Emission factors of wood and charcoal-fired cookstoves. Biomass Bioenerg. 2002;23:453–469. http://dx.doi.org/10.1016/S0961-9534(02)00072-7

33. Zhang J, Smith KR, Uma R, Ma Y, Kishore VVN, Lata K, et al. Carbon monoxide from cookstoves in developing countries 1: Emission factors. Chemosphere. 1999;1:353–366. http://dx.doi.org/10.1016/S1465-9972(99)00004-5

34. Masekameni D, Makonese T, Masekameni D, Forbes P, Annegarn HJ. A comparison of emissions and thermal efficiency of three improved liquid fuel stoves. In: Proceedings of the Domestic Use of Energy (DUE) Conference; 2015 March 31 – April 02; Cape Town, South Africa. Cape Town: IEEE Conference Publications; 2015. p. 71–76. http://dx.doi.org/10.1109/DUE.2015.7102965

35. Language B, Piketh SJ, Burger RP. Correcting respirable photometric particulate measurements using a gravimetric sampling method. Clean Air J. 2016;26(1):10–15. http://dx.doi.org/10.17159/2410-972X/2016/v26n1a7

36. Kimemia D, Makonese T, Annegarn HJ. Comparative evaluation of emissions from selected paraffin lamps and a paraffin thermoelectric generator. In: Proceedings of the Domestic Use of Energy (DUE) Conference; 2015 March 31 – April 02; Cape Town, South Africa. Cape Town: IEEE Conference Publications; 2015. p. 57–62. http://dx.doi.org/10.1109/DUE.2015.7102963

37. Liu GZ, Swor TA, Schauer JJ, Debilzen JA, Severance CL. A source dilution sampling system for characterization of engine emissions under transient or steady-state operation. Aerosol Sci Tech. 2008;42:270–280. http://dx.doi.org/10.1080/02786820801992907

AUTHORS:
Mari Molvik[1,2]
Elin Helland[1]
Siphosenkosi G. Zulu[3]
Elisabeth Kleppa[2,4]
Kristine Lillebo[2]
Svein G. Gundersen[5,6]
Jane D. Kvalsvig[3]
Myra Taylor[3]
Eyrun F. Kjetland[2,3]
Birgitte J. Vennervald[1]

AFFILIATIONS:
[1]Parasitology and Aquatic Diseases, University of Copenhagen, Copenhagen, Denmark
[2]Norwegian Centre for Imported and Tropical Diseases, Department of Infectious Diseases Ullevaal, Oslo University Hospital, Oslo, Norway
[3]Discipline of Public Health Medicine, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa
[4]Faculty of Medicine, University of Oslo, Oslo, Norway
[5]Research Department, Sorlandet Hospital, Kristiansand, Norway
[6]Institute for Global Development and Planning, University of Agder, Kristiansand, Norway

CORRESPONDENCE TO:
Mari Molvik

EMAIL:
marimolvik@hotmail.com

DATES:
Received: 25 Aug. 2016
Revised: 17 Oct. 2016
Accepted: 19 Oct. 2016

KEYWORDS:
urogenital schistosomiasis; KwaZulu-Natal; *Ascaris lumbricoides; Trichuris trichiura;* schoolchildren

HOW TO CITE:
Molvik M, Helland E, Zulu SG, Kleppa E, Lillebo K, Gundersen SG, et al. Co-infection with *Schistosoma haematobium* and soil-transmitted helminths in rural South Africa. S Afr J Sci. 2017;113(3/4), Art. #2016-0251, 6 pages. http://dx.doi.org/10.17159/sajs.2017/20160251

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Norwegian Centre for Imported and Tropical Diseases, Oslo University Hospital Ullevaal (VIRUUS); S.G. Sønneland Foundation; Shipowner Tom Wilhelmsen Foundation

Co-infection with *Schistosoma haematobium* and soil-transmitted helminths in rural South Africa

Schistosomiasis and soil-transmitted helminthiasis are among the most prevalent neglected tropical diseases and may lead to severe consequences. We assessed the extent of co-infection between *Schistosoma haematobium* and the soil-transmitted helminths (STHs) *Ascaris lumbricoides* and *Trichuris trichiura* in schoolgirls in the rural areas of KwaZulu-Natal, South Africa. We also explored if *S. haematobium* can serve as a predictor for soil-transmitted helminths in this area. From 15 selected schools, 726 primary schoolgirls aged 10–12 years provided both urine and stool samples. The samples were examined for the presence of eggs using the urine sedimentation technique for *S. haematobium* and the Kato Katz technique for STHs. Pearson's chi-square test was used to calculate the association and Spearman's rank correlation was used for the correlation analysis. There was a highly significant correlation between *S. haematobium* and STHs at a school level (Spearman's correlation coefficient =0.93; $p<0.001$). The prevalences were found to be 36.9% and 38.8% for *S. haematobium* and STHs, respectively. A significant association was found between *S. haematobium* and STHs (odds ratio =2.05; confidence interval =1.58–2.93; $p<0.001$). Indirect indicators of urogenital schistosomiasis (e.g. water contact and haematuria) were significantly associated with *A. lumbricoides* and *T. trichiura* infection. We have demonstrated a highly significant correlation and overall association between urogenital schistosomiasis and *A. lumbricoides* and *T. trichiura*. We cautiously suggest that all *S. haematobium* endemic areas should be treated for STH infections.

Significance:

- The prevalences of urogenital schistosomiasis and soil-transmitted helminth infections were highly significantly correlated.
- More than half (60%) of the investigated schools are in need of annual treatment for *S. haematobium* infection.
- Almost half of the infected schoolgirls had a heavy intensity of *S. haematobium* infection.
- Nearly all the schools investigated require treatment for soil-transmitted helminthiasis once or even twice per year.
- This study can contribute to the epidemiological planning process of the deworming programme.

Introduction

Schistosomiasis and soil-transmitted helminthiasis represent the most common neglected tropical diseases and may cause acute and chronic illness.[1] The most prevalent schistosome species in South Africa is *Schistosoma haematobium*, which causes urogenital schistosomiasis.[2] It is estimated that 5.2 million people in South Africa are infected with *S. haematobium*.[3] The total number of people infected with soil-transmitted helminths (STHs) in South Africa is unknown, but according to the World Health Organization (WHO) approximately 3.2 million children require treatment in South Africa.[4] These helminth infections may have serious consequences in children, which could lead to decreased growth and stunting, decreased cognitive development and school performance and increased school absenteeism.[5,6] Furthermore, infection with STHs is associated with anaemia and malnutrition[5,7] and urogenital schistosomiasis may also lead to anaemia, dysuria, haematuria, infertility and, in some instances, bladder cancer[8] and may increase the risk of HIV infection in women[6,9,10]. The diseases may be controlled by periodic treatment with so-called preventive chemotherapy.[11]

Studies have demonstrated that deworming programmes significantly improve learning, growth and school attendance among schoolchildren[1,12] and mass treatment programmes integrated with other health services have been found to be the most cost-effective approach to combat the helminth infections[11].

Previous studies have shown that areas endemic for schistosomiasis are often also endemic for soil-transmitted helminthiasis.[13] *S. haematobium* is transmitted to humans through infested fresh water and STHs are transmitted from contaminated soil in areas that lack adequate sanitation, as eggs are excreted in human urine and faeces, respectively.[1] Improvements in water, sanitation and hygiene are important to prevent all helminth infections[14] and in South Africa, sanitation programmes have been rolled out but have not yet reached all areas[15]. In KwaZulu-Natal, 14% of households have never had access to potable water.[15] Furthermore, even people who have access to taps providing potable water must often use the river because of irregular water supply and/or long queues for drinking water at the communal taps.[16]

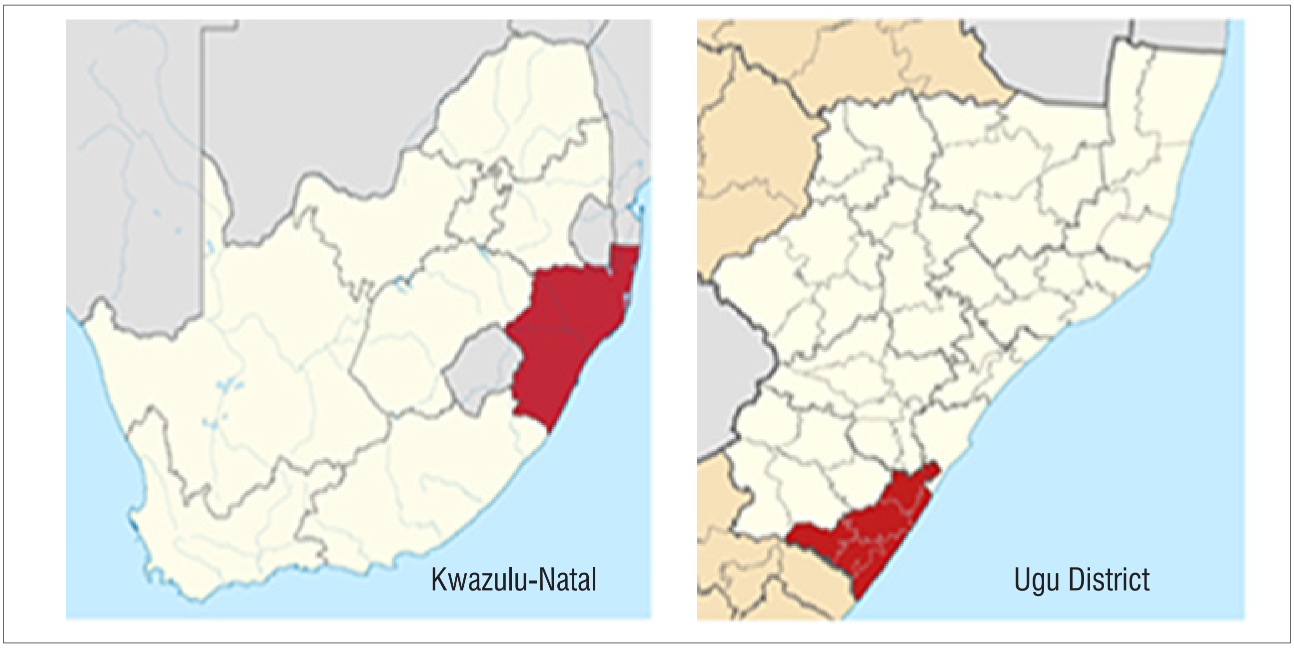


Figure 1: Map showing KwaZulu-Natal Province and Ugu District within KwaZulu-Natal, South Africa.

Soil-transmitted helminthiasis is most commonly diagnosed by stool collection and microscopy, requiring willing patients and trained laboratory staff.[1] *S. haematobium* prevalence may be assessed in children by detecting macro- or microscopic haematuria through a questionnaire or by urine dipstick. The latter technique has demonstrated to be quick, easy to perform, and highly sensitive and specific in children[17-19], although final diagnosis must be confirmed by microscopy or more advanced methods[20].

The Integrated School Health Programme's policy in South Africa recommends treatment for all children in areas endemic for schistosomiasis and soil-transmitted helminthiasis, but this recommendation has still to be implemented[21]; and these areas have still to be identified[22]. In order to make decisions on where to treat, countrywide mapping of helminth infections is needed.

In this study, we aimed to explore the burden of water- and soil-transmitted helminth infections in a middle-income country like South Africa and to determine if infection with *S. haematobium* could serve as a predictor for STHs.

Methods

Study area

The study was conducted between September 2009 and November 2010 in Ugu District in the KwaZulu-Natal Province of South Africa (Figure 1). The study site is endemic for *S. haematobium* and STHs.[23-26] The total population of the District is 709 918; over 50% are female and 33% are under the age of 14.[27]

Study population

Of the 309 primary schools in the area, 18 schools in rural Ugu were randomly selected for inclusion. All girls between the ages of 10 and 12 were invited to participate, and a total of 1057 girls participated in the study. In three of the schools, fewer than 10 girls provided urine and stool samples, and the schools were therefore excluded from the calculations. From the 15 included schools, 726 individuals submitted both stool and urine samples. Girls who did not provide at least one urine and one stool sample were excluded. No mass treatment had been conducted in this cohort of children before the data collection.

Parasitological examination

Urine specimens were collected for three consecutive days between 10:00 and 14:00. Samples were transported to the laboratory in dark cooler boxes. After arrival, 1 mL of 2% tincture of merthiolate in 5% formalin solution was added to each 10-mL urine sample for preservation. Two samples from each participant (labelled A and B) were registered daily (six samples in total per girl). Within the same week, the samples were centrifuged and microscopically investigated for *S. haematobium* eggs. The egg counts were based on the 10 mL of urine and each slide was read independently by two technicians. One stool sample was provided by each girl and the stool samples were kept at 4 °C until they were processed within 24 h after collection. The Kato-Katz technique was used for diagnosing STHs.[28,29] The stool sample was divided into two (A and B) and each sample of 41.9 mg was prepared on a slide and investigated by microscope by two laboratory technicians. Because of the duration between sampling and preservation, it was not possible to identify hookworm eggs in the stool samples.

Ethical considerations and treatment

Three ethics committees granted permission to undertake this study. The Biomedical Research Ethics Committee of the University of KwaZulu-Natal (reference BF029/07), the KwaZulu-Natal Department of Health (Pietermaritzburg, 3 February 2009, reference HRKM010-08) and the Regional Norwegian Ethics Committee gave ethical clearance (reference 469-07066a1.2007.535). Further, both the Departments of Health and Basic Education in Ugu District gave permission. Prior to the study, information meetings were held for the parents, principals, school governing bodies and teachers. Assent was given by each girl, and informed consent forms were signed by parents/guardians. Treatment with a single dose of praziquantel (40 mg/kg) for schistosomiasis was offered to all by the Department of Health and information about possible side effects was given. Treatment for STH was offered at the local clinics.

Data analysis

If at least one egg was found in the urine or stool sample, the person was registered as positive. Schools were categorised into risk groups according to WHO guidelines.[1] Intensity of *S. haematobium* was expressed as eggs per 10 mL of urine, based on the maximum egg count of the three urine samples provided. 'Light infection' is defined as 1–50 eggs per 10 mL of urine and 'heavy infection' as more than 50 eggs per 10 mL of urine.[1] The median egg count of the study population was calculated for *S. haematobium*. The egg counts for the STHs were stopped at 500 eggs per slide and so it was not possible to calculate median or intensity of egg excretion in the high excretors and therefore for the STHs.

Interviews

Research assistants interviewed each girl in the local language (isiZulu) using a pre-designed questionnaire, the results of which are described elsewhere.[25] Interviews included questions on water contact and urogenital symptoms such as pain when urinating and having observed red urine.

Statistical analysis

Data were entered into MS Excel spreadsheets and exported to SPSS Statistics (IBM SPSS Version 22). When the data did not have a normal distribution, a non-parametric statistical test was used. A p-value lower than 0.05 was considered statistically significant. Pearson's chi-square test and odds ratios were used to calculate the association between the helminth species. Pearson's chi-square and Mann–Whitney U tests were used for comparison of age or household size and categorical data of included and excluded cases. The helminth prevalences found in each school were compared using Spearman's rank correlation.

Results

Table 1 shows the descriptive characteristics of the study population. Over a third (37%; 268/726) of the girls was infected with schistosomiasis and the prevalence of STHs (either *A. lumbricoides* or *T. trichiura*) was 38.8% (282/726). The egg counts for *S. haematobium* ranged from 1 to 624 eggs per 10 mL (with a median of 21). Heavy intensity of *S. haematobium* was found in 47.8% (128/268) of the girls positive for this infection. Infection with *S. haematobium* was not only significantly associated with STH infection (either ascariasis or trichuriasis) (odds ratio =2.15; 95% confidence interval =1.58–2.93; $p<0.001$), but also with each of the species (Table 1). Red urine, dysuria, washing blankets in the river and swimming in the river were associated with *S. haematobium* (data not shown). These variables were also investigated as a predictor for STHs and found to be significantly associated (Table 1). Age did not influence any of the associations. The included ($n=726$) and excluded ($n=297$) cases had similar exposures to risk water and similar family structures as those shown in Table 2.

School-level analysis

The prevalence of *S. haematobium* found in each investigated school ranged between 8.8% and 70.0% (median 50.0%) and the prevalence of STHs in the schools ranged between 0 and 63.0% (median 37.5%).

Table 1: Descriptive characteristics of the study population, prevalence of *Schistosoma haematobium, Ascaris lumbricoides, Trichuris trichiura,* and association with soil-transmitted helminth (STH) infection

Category	Prevalence	Association with STH infection
Schistosoma haematobium	36.9% (268/726)	OR = 2.15, 95% CI = 1.58–2.93, $p<0.001$
Ascaris lumbricoides	25.6% (184/726)	OR = 1.62, 95% CI = 1.15–2.28, $p=0.005$
Trichuris trichiura	28.0% (203/726)	OR = 2.33, 95% CI = 1.67–3.24, $p<0.001$
Access to water from communal tap (mean)	73.3% (530/723)	No association
Have an indoor tap	22.3% (161/723)	No association
Use river as a source for drinking water	1.7% (12/723)	No association
Swim in river/dam	32.4% (235/726)	Adjusted OR = 1.38, 95% CI = 1.00–1.89, $p=0.05$
Wash blankets in river	26.2% (190/726)	Adjusted OR = 1.80, 95% CI = 1.28–2.52, $p<0.001$
Red urine	17.6% (127/723)	Adjusted OR = 1.44, 95% CI = 1.14–1.83, $p=0.002$
Dysuria	21.5% (155/721)	Adjusted OR = 1.55, 95% CI = 1.19–2.03, $p=0.001$

OR, odds ratio; CI, confidence interval

Table 2: Comparison of characteristics of included cases (stool, urine and questionnaire provided) and excluded cases

Category	Included (n=726)	Excluded (n=297)	p-value
Median age in years (range)	11 (9–13)	11 (9–13)	0.98‡
Median household size	6 (2–19)	6 (2–19)	0.51‡
Access to piped/potable water (mean)	97.0% (700/722)	96.6% (287/297)	0.79†
Swim in river/dam	32.4% (235/726)	29.6% (98/331)	0.37†
Ever lived in city	6.6% (47/710)	7.4% (22/297)	0.65†
Median employed people in household	1 (0–8)	1 (0–9)	0.66‡
Mother as main caregiver	65.3% (473/724)	64.5% (193/299)	0.81†
Mothers highest education is high school or higher	63.3% (386/610)	64.4% (159/247)	0.76†

‡Mann–Whitney U test, †chi-square test

Figure 2 shows the correlation between *S. haematobium* and STHs based on school prevalences. The correlation coefficient was found to be 0.93 ($p<0.001$). Of the investigated schools in Ugu District, 60% (9/15) were found to have a prevalence of urogenital schistosomiasis above 50% and the same applied for 20% (3/15) of the schools for soil-transmitted helminthiasis. Just above 70% (11/15) of the schools had a prevalence of STH infection between 20% and 50%.

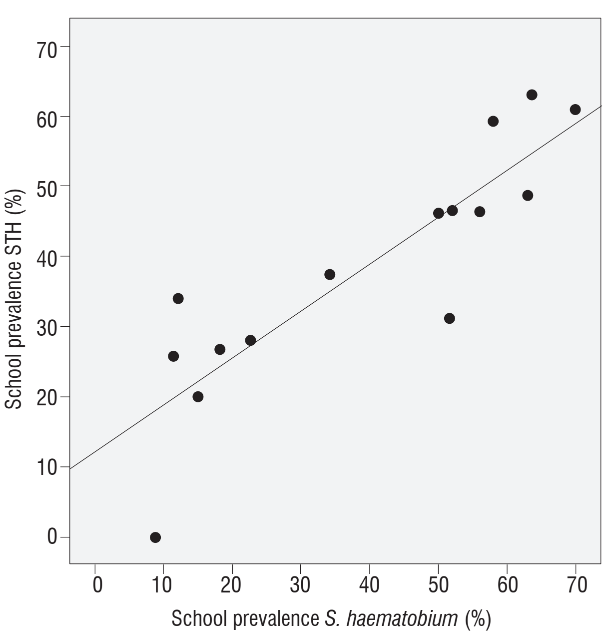


Figure 2: Significant correlation (Spearman's correlation coefficient =0.93; $p<0.001$) between prevalence of *Schistosoma haematobium* and soil-transmitted helminths (*Ascaris lumbricoides* and *Trichuris trichiura*) in schoolchildren.

Discussion

At a school level, the prevalence of urogenital schistosomiasis and STH infections were highly significantly correlated. Overall prevalence of infection with *S. haematobium* and STHs was significantly associated. In addition, indirect indicators of urogenital schistosomiasis were significantly associated with *A. lumbricoides* and *T. trichiura* infection. Based on the WHO recommendations, this study shows that 60% of the investigated schools are in need of annual treatment for *S. haematobium* because they show a prevalence in excess of 50%.[1] For soil-transmitted helminthiasis, the recommendation is treatment once per year if the school is classified as low risk (a prevalence of 20–50%) and twice a year if classified as high risk (a prevalence greater than 50%).[1] Nearly all the schools investigated require treatment for soil-transmitted helminthiasis once or even twice per year. Almost half of the infected school girls showed a heavy intensity of *S. haematobium* infection.

The observed associations are consistent with previous studies[13,30] and contribute to the evidence showing that co-infection occurs in areas endemic for *S. haematobium*. This present study shows that the prevalence of *S. haematobium*, *A. lumbricoides* and *T. trichiura* was within the same range as reported previously in the area.[23,24,31] Our study only included participants who provided both urine and stool samples, but there was no difference between the included and excluded girls in terms of household size, water contact and age.

Intensity of infection has been shown to be positively correlated with morbidity, and measuring intensity allows quantification of the proportion of individuals with high intensity of infection who may suffer serious consequences.[1] However, previous studies have shown that the majority of infected children harbour light infections and relatively few people suffer severe morbidity.[32] The WHO recommends calculating the intensity of infection at community level to estimate the morbidity rate in the same area.[1] This calculation can be done by estimation of the mean intensity, or more comprehensively by classes of intensity,[33] as done in this paper. Our results showed that almost 50% of the girls had a heavy intensity *S. haematobium* infection (egg counts above 500). We did not investigate the intensity of STH infection. However, persons with a heavy intensity of one helminth species have been found to be more likely to harbour a heavy intensity of the other helminth species.[34] It has been hypothesised that overdispersion may be explained by a predisposition of some individuals to helminth infections, although the underlying reasons remain poorly understood.[32,35] Possible factors are differences in susceptibility and immunity against infection.[32] Another possibility is that because of poor sanitation in general, infestation in only a few latrines can lead to contamination of the water and soil by these helminths, as there is often limited access to clean water. Jinabhai et al.[31] reported this pattern of intensity for *S. haematobium* and *T. trichiura* in southern KwaZulu-Natal.

Saathoff et al.[36] found that amongst 10- to 12-year-old pupils, the boys had about 10% lower prevalence of *S. haematobium* than the girls and the prevalence of *A. lumbricoides*, but not *T. trichiura*, also differed significantly between boys and girls. On the other hand, the opposite tendency was found among schoolchildren in Zanzibar: boys were slightly more affected by *S. haematobium*, both regarding prevalence and intensity, and significantly more heavily infected with *T. trichiura* than the girls.[30] Hence, the results from this study on schoolgirls cannot be extrapolated to different populations because local gender-related habits can differ.

Low socio-economic status is intimately related to the presence of helminth infections.[37] Members of poorer families are found to be more frequently infected with *A. lumbricoides* and *T. trichiura* and more likely to have both schistosomiasis and soil-transmitted helminthiasis.[37] Poor economic growth has been shown to maintain high prevalences of STH infections.[38] However, the connection between poverty and STHs is complex, because STH prevalence in turn is a contributing factor to poor economic growth and the anaemia caused by soil-transmitted helminthiasis is associated with reduced work output.[38] It has been shown that children from poorer families are more likely to drop out of school.[39] There is an 85% school enrolment in South Africa.[40] This study only included enrolled schoolgirls and those present at the days of sample collections. Schoolgirls may be absent as a result of sickness caused by urogenital schistosomiasis or soil-transmitted helminthiasis or other health issues, or may stay home for family reasons. Hence, these prevalences and intensities of infection may be underestimated.

Different activities involving water contact are highly associated with *S. haematobium* infection,[41] and washing hands with soap after defecation and before eating is protective against STHs (except hookworm) by limiting the faeco-oral route of infection.[37] Access to clean fresh water has been associated with a significant reduction in STH infection.[30] Furthermore, not using a tap as the source for drinking water is positively associated with urogenital schistosomiasis.[37] The Ugu District Municipality Integrated Development Plan confirmed that constructions of sanitary facilities and water sources have taken place in the area in recent years, but improvements have not reached all areas.[27] The prevalence of the helminth infections in the schools varied from 8.8% to 70.0% for *S. haematobium* and 0 to 63% for STHs. In order to explain the differences between schools, more background information about the socio-economic status of the participants and their community, their access to water and sanitation facilities and local risk factors is needed.[14,37,39,41]

Because of both time and resource constraints, only one faecal sample was collected from each participant in this study, although three faecal samples on three consecutive days markedly raises the sensitivity.[29,42] Because egg excretion has a day-to-day variation, there may have been an under- or overestimation of both prevalence and intensity of infection.[42] Hookworm was not investigated, because of logistical constraints, although it would have been desirable as part of the total examination for STHs.

The difference in transmission patterns of *S. haematobium* and STHs is essential when it comes to prevention and control. Although the helminth infections have many of the same risk factors, the transmission cycle of *S. haematobium* requires fresh water with snails as the intermediate host.[2] The distribution and transmission of schistosomiasis is therefore known to be highly focal – depending on different factors such as presence of water and local environmental conditions – whereas STHs are more widely distributed.[1] In the southern part of South Africa, schistosomiasis is not found more than 300 m above sea level, and often not in urban areas. STHs, however, are found in the Cape Peninsula and in urban slums.[43] However, because of the great burden of helminth infections, WHO recommends coordinated interventions to secure a joint and synergic control of the diseases.[11] South Africa's first school-based helminth control programme in KwaZulu-Natal, from 1997 to 2000[23], showed a decrease in prevalence for all the helminths investigated (*S. haematobium, A. lumbricoides, T. trichiura* and hookworms) after treating the schoolchildren with chemotherapy.[44] Despite the successful results, the programme was discontinued and it has been suggested that this discontinuation was a result of a shift in funding priorities consequent to the massive burden of HIV and tuberculosis.[43] Newer studies have shown that patients with urogenital schistosomiasis may be at higher risk for HIV acquisition, which should be considered of importance for control of the helminth infection.[45] The South African health authorities have recently launched an initiative to combat helminth infections in endemic areas and this study can contribute to the epidemiological planning process of the deworming programme.

Considering the significant correlation and association between urogenital schistosomiasis and STH infection shown in this study, using *S. haematobium* as a predictor for STHs, at least in rural parts of districts where urogenital schistosomiasis is endemic, may facilitate the processes of planning interventions.[1] Stool sampling and analyses are laborious and the faecal collection may be challenging as a result of age and cultural differences of the study participants.[1] Despite the fact that the investigated helminth infections have different transmission patterns and require different treatment interventions, this study shows that simple urine analyses, or possibly even water contact information, may indicate which persons and rural schools need intervention for both urogenital schistosomiasis and soil-transmitted helminthiasis. However, because of the focal nature of schistosomiasis transmission, ultimately, both stool and urine prevalence surveys should be done countrywide with a view of mapping disease distribution.

Acknowledgements

We warmly thank the schoolgirls who participated in this study and their carers. We are grateful for the help from the BRIGHT clinical staff for sample collection and laboratory work – this study could not have been done without their hard work and effort. We thank R.F. Manyaira for able assistance, S.D. Holmen for SPSS support and P.D. Ndhlovu, H.N. Galappaththi-Arachchige and I. Lindstroem for valuable feedback. We also thank the Norwegian Centre for Imported and Tropical Diseases, Oslo University Hospital Ullevaal (VIRUUS), Shipowner Tom Wilhelmsen Foundation, S.G. Sønneland Foundation and General Travel Scholarships for funding. The funders had no role in data collection, study design and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

M.M. and E.H. worked on the original report, analysed the data and conceptualised the paper together with E.F.K. and B.J.V.; S.G.Z. and E.F.K. contributed to data collection. All authors contributed to the editing of the manuscript.

References

1. World Health Organization. Helminth control in school-age children. A guide for managers of control programmes. Geneva: Preventive Chemotherapy and Transmission Control (PCT), Department of Control of Neglected Tropical Diseases (NTD); 2011. Available from: http://apps.who.int/iris/bitstream/10665/44671/1/9789241548267_eng.pdf

2. Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. Lancet. 2006;368(9541):1106–1118. http://dx.doi.org/10.1016/S0140-6736(06)69440-3

3. World Health Organization. Neglected tropical diseases: PCT databank: Schistosomiasis [database on the Internet]. c2014 [cited 2014 Nov 03]. Available from: http://www.who.int/neglected_diseases/preventive_chemotherapy/sch/en/

4. World Health Organization. Neglected tropical diseases: PCT databank: Soil-transmitted helmithiases [database on the Internet]. c2014 [cited 2014 Nov 11]. Available from: http://www.who.int/neglected_diseases/preventive_chemotherapy/sth/en/

5. Stephenson LS, Latham MC, Ottesen EA. Malnutrition and parasitic helminth infections. Parasitology. 2000;121:23–38. http://dx.doi.org/10.1017/S0031182000006491

6. King CH, Dangerfield-Cha M. The unacknowledged impact of chronic schistosomiasis. Chronic Illn. 2008;4(1):65–79. http://dx.doi.org/10.1177/1742395307084407

7. Ezeamama AE, Friedman JF, Olveda RM, Acosta LP, Kurtis JD, Mor V, et al. Functional significance of low-intensity polyparasite helminth infections in anemia. J Infect Dis. 2005;192(12):2160–2170. http://dx.doi.org/10.1086/498219

8. Vennervald BJ, Polman K. Helminths and malignancy. Parasite Immunol. 2009;31(11):686–696. http://dx.doi.org/10.1111/j.1365-3024.2009.01163.x

9. Kjetland EF, Ndhlovu PD, Gomo E, Mduluza T, Midzi N, Gwanzura L, et al. Association between genital schistosomiasis and HIV in rural Zimbabwean women. AIDS. 2006;20(4):593–600. http://dx.doi.org/10.1097/01.aids.0000210614.45212.0a

10. Downs JA, Mguta C, Kaatano GM, Mitchell KB, Bang H, Simplice H, et al. Urogenital schistosomiasis in women of reproductive age in Tanzania's Lake Victoria region. Am J Trop Med Hyg. 2011;84(3):364–369. http://dx.doi.org/10.4269/ajtmh.2011.10-0585

11. World Health Organization. Preventive chemotherapy in human helminthiasis. Coordinated use of anthelminthic drugs in control interventions: A manual for health professionals and programme managers. Geneva: Preventive Chemotherapy and Transmission Control (PCT), Department of Control of Neglected Tropical Diseases (NTD); 2006. Available from: http://apps.who.int/iris/bitstream/10665/43545/1/9241547103_eng.pdf

12. Dickson R, Awasthi S, Williamson P, Demellweek C, Garner P. Effects of treatment for intestinal helminth infection on growth and cognitive performance in children: Systematic review of randomised trials. BMJ. 2000;320(7251):1697–1701. http://dx.doi.org/10.1136/bmj.320.7251.1697

13. Yajima A, Gabrielli AF, Montresor A, Engels D. Moderate and high endemicity of schistosomiasis is a predictor of the endemicity of soil-transmitted helminthiasis: A systematic review. Trans R Soc Trop Med Hyg. 2011;105(2):68–73. http://dx.doi.org/10.1016/j.trstmh.2010.11.003

14. Campbell SJ, Savage GB, Gray DJ, Atkinson JA, Soares Magalhaes RJ, Nery SV, et al. Water, sanitation, and hygiene (WASH): A critical component for sustainable soil-transmitted helminth and schistosomiasis control. PLoS Negl Trop Dis. 2014;8(4), Art. #e2651, 5 pages. http://dx.doi.org/10.1371/journal.pntd.0002651

15. South African Human Rights Commission (SAHRC). Report on the right to access sufficient water and decent sanitation in South Africa. Johannesburg: SAHRC; 2014. Available from: https://www.sahrc.org.za/home/21/files/FINAL%204th%20Proof%204%20March%20-%20Water%20%20Sanitation%20low%20res%20(2).pdf

16. Ugu District Municipality. Water and sanitation services: Ugu District [homepage on the Internet]. c2014 [cited 2014 Nov 04]. Available from: http://www.ugu.gov.za/Water_Services.aspx.

17. Savioli L, Dixon H, Kisumku UM, Mott KE. Control of morbidity due to *Schistosoma haematobium* on Pemba Island; selective population chemotherapy of schoolchildren with haematuria to identify high-risk localities. Trans R Soc Trop Med Hyg. 1989;83(6):805–810. http://dx.doi.org/10.1016/0035-9203(89)90336-2

18. Houmsou RS, Kela SL, Suleiman MM. Performance of microhaematuria and proteinuria as measured by urine reagent strips in estimating intensity and prevalence of *Schistosoma haematobium* infection in Nigeria. Asian Pac J Trop Med. 2011;4(12):997–1000. http://dx.doi.org/10.1016/S1995-7645(11)60233-2

19. Brooker S, Kabatereine NB, Gyapong JO, Stothard JR, Utzinger J. Rapid mapping of schistosomiasis and other neglected tropical diseases in the context of integrated control programmes in Africa. Parasitology. 2009;136(13):1707–1718. http://dx.doi.org/10.1017/S0031182009005940

20. Gundersen SG, Kjetland EF, Poggensee G, Helling-Giese G, Richter J, Chitsulo L, et al. Urine reagent strips for diagnosis of schistosomiasis haematobium in women of fertile age. Acta Trop. 1996;62(4):281–287. http://dx.doi.org/10.1016/S0001-706X(96)00029-0

21. Department of Basic Education (DBE) and Department of Health (DoH). Integrated school health policy 2012. Pretoria: DBE/DoH; 2012. Available from: http://www.education.gov.za/Portals/0/Documents/Policies/INTEGRATED%20SCHOOL%20HEALTH%20POLICYB-W_1.pdf?ver=2014-06-14-172322-000

22. Mkhize-Kwitshana ZL, Mabaso MLH. The neglected triple disease burden and interaction of helminths, HIV and tuberculosis: An opportunity for integrated action in South Africa. S Afr Med J. 2014;104(4):258–259. http://dx.doi.org/10.7196/SAMJ.7947

23. Appleton CC, Kvalsvig JD. A school-based helminth control programme successfully implemented in KwaZulu-Natal. South Afr J Epidemiol Infect. 2006;21(2):55–67.

24. Appleton CC, Maurihungirire M, Gouws E. The distribution of helminth infections along the coastal plain of Kwazulu-Natal province, South Africa. Ann Trop Med Parasitol. 1999;93(8):859–868.

25. Hegertun IEA, Sulheim Gundersen KM, Kleppa E, Zulu SG, Gundersen SG, Taylor M, et al. *S. haematobium* as a common cause of genital morbidity in girls: A cross-sectional study of children in South Africa. PLoS Negl Trop Dis. 2013;7(3), Art. #e2104, 8 pages. http://dx.doi.org/10.1371/journal.pntd.0002104

26. Jinabhai CC, Taylor M, Coutsoudis A, Coovadia HM, Tomkins AM, Sullivan KR. Epidemiology of helminth infections: Implications for parasite control programmes, a South African perspective. Public Health Nutr. 2001;4(6):1211–1219. http://dx.doi.org/10.1079/PHN2001180

27. Ugu District Municipality. Ugu District Municipality integrated development plan review 2013/2014. Port Shepstone: Ugu District Municipality; 2013.

28. Katz N, Chaves A, Pellegrino J. A simple device for quantitative stool thick-smear technique in schistosomiasis mansoni. Rev Inst Med Trop Sao Paulo. 1972;14(6):397–400.

29. World Health Organization (WHO). Basic laboratory methods in medical parasitology. Geneva: WHO; 1991. Available from: http://apps.who.int/iris/bitstream/10665/40793/1/9241544104_%28part1%29.pdf

30. Stothard JR, French MD, Khamis IS, Basanez MG, Rollinson D. The epidemiology and control of urinary schistosomiasis and soil-transmitted helminthiasis in schoolchildren on Unguja Island, Zanzibar. Trans R Soc Trop Med Hyg. 2009;103(10):1031–1044. http://dx.doi.org/10.1016/j.trstmh.2009.03.024

31. Jinabhai CC, Taylor M, Coutsoudis A, Coovadia HM, Tomkins AM, Sullivan KR. A health and nutritional profile of rural school children in KwaZulu-Natal, South Africa. Ann Trop Paediatr. 2001;21(1):50–58. http://dx.doi.org/10.1080/02724930020028920

32. Hotez PJ, Bundy DAP, Beegle K, Brooker S, Drake L, De Silva N, et al. Helminth infections: Soil-transmitted helminth infections and schistosomiasis. In: Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB, et al, editors. Disease control priorities in developing countries. 2nd ed. Washington DC: World Bank; 2006.

33. Montresor A. Arithmetic or geometric means of eggs per gram are not appropriate indicators to estimate the impact of control measures in helminth infections. Trans R Soc Trop Med Hyg. 2007;101(8):773–776. http://dx.doi.org/10.1016/j.trstmh.2007.04.008

34. Brooker S, Miguel EA, Moulin S, Luoba AI, Bundy DA, Kremer M. Epidemiology of single and multiple species of helminth infections among school children in Busia District, Kenya. East Afr Med J. 2000;77(3):157–161. http://dx.doi.org/10.4314/eamj.v77i3.46613

35. Bundy DA, Cooper ES, Thompson DE, Didier JM, Anderson RM, Simmons I. Predisposition to *Trichuris trichiura* infection in humans. Epidemiol Infect. 1987;98(1):65–71. http://dx.doi.org/10.1017/S0950268800061719

36. Saathoff E, Olsen A, Magnussen P, Kvalsvig JD, Becker W, Appleton CC. Patterns of *Schistosoma haematobium* infection, impact of praziquantel treatment and re-infection after treatment in a cohort of schoolchildren from rural KwaZulu-Natal/South Africa. BMC Infect Dis. 2004;4:40. http:dx.doi.org/10.1186/1471-2334-4-40

37. Balen J, Raso G, Li YS, Zhao ZY, Yuan LP, Williams GM, et al. Risk factors for helminth infections in a rural and a peri-urban setting of the Dongting Lake area, People's Republic of China. Int J Parasitol. 2011;41(11):1165–1173. http://dx.doi.org/10.1016/j.ijpara.2011.07.006

38. De Silva NR, Brooker S, Hotez PJ, Montresor A, Engels D, Savioli L. Soil-transmitted helminth infections: Updating the global picture. Trends Parasitol. 2003;19(12):547–551. http://dx.doi.org/10.1016/j.pt.2003.10.002

39. Bock J. Evolutionary demography and intrahousehold time allocation: School attendance and child labor among the Okavango Delta Peoples of Botswana. Am J Hum Biol. 2002;14(2):206–221. http://dx.doi.org/10.1002/ajhb.10040

40. The World Bank. Net enrolment rate, primary, both sexes (%) [homepage on the Internet]. c2012 [cited 2014 Dec 16]. Available from: http://data.worldbank.org/indicator/SE.PRM.NENR/countries/ZA-ZF-XT?display=graph

41. Rudge JW, Stothard JR, Basanez MG, Mgeni AF, Khamis IS, Khamis AN, et al. Micro-epidemiology of urinary schistosomiasis in Zanzibar: Local risk factors associated with distribution of infections among schoolchildren and relevance for control. Acta Trop. 2008;105(1):45–54. http://dx.doi.org/10.1016/j.actatropica.2007.09.006

42. Knopp S, Mgeni AF, Khamis IS, Steinmann P, Stothard JR, Rollinson D, et al. Diagnosis of soil-transmitted helminths in the era of preventive chemotherapy: Effect of multiple stool sampling and use of different diagnostic techniques. PLoS Negl Trop Dis. 2008;2(11), Art. #e331, 8 pages. http://dx.doi.org/10.1371/journal.pntd.0000331

43. Mkhize-Kwitshana ZL, Mabaso MH. Status of medical parasitology in South Africa: New challenges and missed opportunities. Trends Parasitol. 2012;28(6):217–219. http://dx.doi.org/10.1016/j.pt.2012.03.005

44. Kvalsvig JD, Appleton CC, Archer C, Mthethwa P, Memela C, Mpanza JT, et al. The KwaZulu-Natal Parasite Control Programme 1998-2000. Durban: Child Development Programme, School of Psychology and School of Life and Environmental Sciences, UKZN; 2001.

45. Kjetland EF, Leutscher PD, Ndhlovu PD. A review of female genital schistosomiasis. Trends Parasitol. 2012;28(2):58–65. http://dx.doi.org/10.1016/j.pt.2011.10.008

Characterisation of *vumba* and *ubumba* clays used for cosmetic purposes

AUTHORS:
Refilwe Morekhure-Mphahlele[1,2]
Walter W. Focke[3]
Wiebke Grote[4]

AFFILIATIONS:
[1]Department of Chemistry, Tshwane University of Technology, Pretoria, South Africa
[2]Department of Chemical Engineering, University of Pretoria, Pretoria, South Africa
[3]Institute of Applied Materials, Department of Chemical Engineering, University of Pretoria, Pretoria, South Africa
[4]Department of Geology, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Walter Focke

EMAIL:
Walter.Focke@up.ac.za

DATES:
Received: 06 Aug. 2016
Revised: 06 Oct. 2016
Accepted: 08 Oct. 2016

KEYWORDS:
quartz; tradition; chemical composition; mineral composition

HOW TO CITE:
Morekhure-Mphahlele R, Focke WW, Grote W. Characterisation of *vumba* and *ubumba* clays used for cosmetic purposes. S Afr J Sci. 2017;113(3/4), Art. #2016-0105, 5 pages. http://dx.doi.org/10.17159/sajs.2017/20160105

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Tshwane University of Technology

Two traditional cosmetic clays bear similar names in different local South African languages: *vumba* (Tshivenda) and *ubumba* (isiZulu). The wet clays are applied topically for cosmetic purposes by the respective indigenous peoples. Six samples from two South African provinces were characterised using X-ray diffraction, X-ray fluorescence spectroscopy, Fourier transform infrared spectroscopy, thermal gravimetric analysis and scanning electron microscopy. It was found that the samples differed widely with respect to mineralogy and chemical composition. This finding raises the possibility that texture characteristics during application on the skin override composition effects. Of concern is the high levels of quartz found in all the samples as it might pose a health hazard; the lowest value for quartz was 11 wt% for *vumba,* while values for *ubumba* ranged from 26 wt% to 85 wt%. All samples contained varying amounts of silicates in the form of smectite, kaolin, chlorite and plagioclase. Minor amounts of anatase and rutile were present in some samples. Three samples also contained goethite. All samples were essentially free from the toxic elements As, Pb, Hg, Cd, Se and Sb. However, they did contain low levels of chromium and heavy metals such as Cu, Zn and Ni. The pH values of *ubumba* slurries were slightly basic, while those of a *vumba* slurry were slightly acidic.

Significance:
- Wide ranges of composition appear to be acceptable.
- The clays do not contain highly toxic or radioactive elements.
- The high levels of quartz present may pose a human health risk.

Introduction

Throughout human history, clays and clayey soils have been used for religious, artistic, cosmetic and therapeutic purposes. For example, the utilisation of ochre, a clay stained by iron oxides, dates back to the Middle Stone Age. South Africa has a rich tradition in this regard. Evidence of human exploitation, also dating back to the Middle Stone Age, has been found at Klein Kliphuis[1], the Blombos Cave[2], Diepkloof Rock Shelter[3] and the Sibudu Cave[4]. The practice of using ochre-like substances for topical cosmetic applications and other purposes persists to the present day. The red clay pastes are known as *letsoku* in Sotho culture and *ibomvu* among Nguni people and are used by both women and men in traditional ceremonies. This and other lighter-coloured clays apparently also serve as sunscreens.[5]

Clays find application as cosmetic products, i.e. applications in which the preparation is placed in contact with the outside of the human body, e.g. the skin, hair and lips.[6] In this context 'clay' can refer to a mineralogical term. However, it may also denote a natural material composed of very fine-grained minerals that show some plasticity when mixed with an appropriate amount of water.[7] The layered structure and colloidal size of the particles constituting the clay give rise to desired rheological characteristics and sorptive abilities, among others.[7] Natural clays typically contain other minerals such as quartz, feldspars, carbonates, sulfates and iron, aluminium or titanium oxides. These minerals may affect the chemical (e.g. stability, purity), physical (e.g. texture, moisture content, particle size) and toxicological requirements specified for each clay application.[7] Particular attention is paid to the presence of silica as there is sufficient evidence to suggest that it is a potential carcinogen. To avoid requirements for labelling that provides safety information, the silica content must be less than 0.1%.[7]

Traditional clays, used among indigenous people in the Eastern Cape Province of South Africa, have been studied extensively.[5,6,8-11] However, there are few reports dealing with the nature and properties of the yellowish-grey clays that are also used primarily for cosmetic purposes. Interestingly, *letsoku* is usually supplied as a dry powder, whereas other clays, colloquially known as *vumba* and *ubumba* in the rural areas, are supplied in a wet form and provide a skin moisturising effect. This communication provides information on the nature and chemical composition of the latter two clays sourced from KwaZulu-Natal and Limpopo Provinces, respectively. This information is relevant for assessing the suitability of these cosmetic clays for more widespread commercial distribution.[12]

Materials and methods

Materials and sample preparation

The wet light brown to grey clay samples were sourced from three different sites in two South African provinces. The clays were supplied by traditional healers and are reportedly sold locally for topical cosmetic applications, although health benefits are also claimed. In some cases, different samples with distinct properties were supplied from the same general deposit. Details of the sampling locations are given in Table 1.

Table 1: Location of the clay deposits sampled

Sample	Location	GPS coordinates
U1, U2, U3	River pond near Mtubatuba (KwaZulu-Natal)	28.4167°S, 32.1833°E
W1, W2	Nongoma (KwaZulu-Natal)	27.8833°S, 31.6333°E
V1	Mbaleni wetlands, Venda (Limpopo)	30.4906°S, 22.9603°E

The wet clay samples were supplied packaged in clear polyethylene bags by the traditional healers who sell the products for topical applications. They were transferred into polyethylene bottles. Portions were dried at 50 °C overnight before being milled into fine powders (<75 μm) in a milling unit fitted with a tungsten carbide vessel.

Characterisation

X-ray diffraction (XRD) measurements were performed at the Department of Geology (University of Pretoria) on a PANalytical X'Pert PRO X-ray diffractometer in *q-q* configuration, equipped with Fe-filtered Co-Ka radiation (λ = 1.789 Å) and an X'Celerator detector and variable divergence and fixed receiving slits. Samples were prepared according to the standardised PANalytical back-loading system, which provides nearly random distribution of the particles. The data were collected in the angular range 5°≤2θ≤90° with a step size of 0.008° 2θ and a 10-s scan step time. The phases were identified using X'Pert Highscore plus software.

The relative phase amounts (weight %) were estimated by the Rietveld method using Autoquan/BGMN software (GE Inspection Technologies; Kleeberg & Bergmann) which employs the fundamental parameter approach. Autoquan combines the analytical potential of BGMN with an 'easy to operate' user surface. In the Rietveld method, an observed data pattern is compared with a calculated pattern. By varying all the parameters the difference between the calculated and observed patterns is minimised by a least squares procedure, until the best possible fit is obtained. The background was fitted by the polynomial order which was determined automatically, depending on the angular range.

X-ray fluorescence spectroscopy (XRF) analyses were performed at the Department of Geology (University of Pretoria) on powders milled to a fine particle size (<75 μm). The moisture content was determined by weighing an accurate mass of approximately 3 g of powder into an alumina crucible. The crucible was then heated in an oven at 100 °C for 2 h. Thereafter the dehydrated samples were roasted at 1000 °C overnight. A sample mass of 1 g of the residue was fused with about 6 g of lithium tetraborate at 1050 °C for metal oxide determination. The trace metals were determined on samples bound with poly(vinyl alcohol) and pressed into powder briquettes.

Scanning electron microscopy (SEM) was performed in the Laboratory for Microscopy and Microanalysis (University of Pretoria), on the powders. Sample powders were lightly sprinkled onto small stubs of double-sided adhesive carbon tape. Compressed air was used to remove any excess powder. The morphology of the carbon-coated powders was studied with a Zeiss Ultra-55 field emission scanning electron microscope fitted with an InLens detector at an acceleration voltage of 1 kV.

Attenuated total reflection Fourier transform infrared (ATR–FTIR) spectra were recorded on an in-house Perkin Elmer Spectrum 100 Series instrument. The spectra represent averages of 32 scans over the wave number range 500–4000 cm^{-1}.

Thermogravimetric analysis (TGA) was performed on an in-house Perkin Elmer TGA 4000 unit. Alumina crucibles were used and the sample size was about 40 mg. Temperature was scanned from 25 °C to 950 °C at 10 °C/min with air flowing at 50 mL/min.

The cation exchange capacity of the clays was determined using the pH-dependent methylene blue halo method adapted from Kahr and Madsen[13].

Results and discussion

The X-ray diffractograms presented in Figure 1 indicate that the materials are complex mixtures of various minerals. The mineral compositions derived from the XRD data are summarised in Table 2. According to the main mineral phases present, samples V1 and U1 are smectitic clays with major admixtures of quartz and minor amounts of kaolin and plagioclase. In contrast, all the other samples are composed primarily of quartz with intermediate amounts of kaolinite and plagioclase. All samples contain minor amounts of the recalcitrant titanium dioxide, i.e. rutile and anatase. Only samples U1, U3 and W1 contained goethite, which also explains their yellowish colour. An unexpected and surprising observation is the high compositional variability of the samples. This variability extends to the samples sourced from effectively the same locations. This finding leads one to speculate that factors other than mere compositional aspects determine the suitability of the clays for their intended purposes. Perhaps the subjective texture experience during skin application and use might be more important. In the wet, as-supplied form, all samples had a smooth creamy feel when rubbed between the thumb and index finger. However, they also featured varying degrees of grittiness. Sample W2 was particularly gritty, probably because of the presence of silica particles.

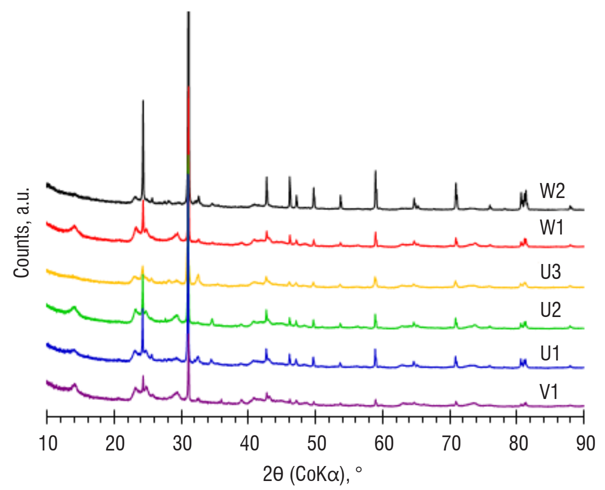


Figure 1: X-ray diffractograms obtained for pulverised clay samples.

Table 2: X-ray diffraction derived mineral content (wt%) estimated using Autoquan software

Mineral	Sample					
	V1	U1	U2	U3	W1	W2
Kaolinite 1:1	7.9±0.9	4.3±0.6	12.8±1.1	10.7±0.8	35.9±1.4	–
Smectite 2:1	67.3±1.6	58.0±1.1	–	–	–	–
Chlorite 2:1:1	5.6±0.9	–	–	–	–	7.4±0.5
Plagioclase	–	5.05±0.75	34.4±1.4	23.7±0.7	–	7.0±1.0
Epidote	6.0±0.6	–	–	–	–	–
Quartz	11.0±0.5	26.1±0.6	47.3±1.3	61.2±1.0	46.9±1.3	84.5±1.0
Anatase	2.2±0.2	1.1±0.1	5.5±0.5	1.1±0.1	5.4±0.5	–
Rutile	–	–	–	–	2.7±0.6	1.2±0.3
Goethite	–	5.5±0.4	–	3.3±0.3	9.0±0.8	–

A significant finding is that all samples contained substantial amounts of quartz. The quartz content varied from 11.4 wt% in sample V1 to 84.5 wt% in sample W2. According to Khiari et al.[14], a quartz content above 15 wt% is cause for concern, even for topical applications, as there is sufficient evidence of quartz carcinogenicity.

The chemical composition results in Table 3 confirm the high silicon content expected in view of the silica and the silicate minerals present in the samples. In particular, it confirms the very high quartz content of sample W2. Sample V1 shows the highest iron content, despite the fact that it apparently contains no goethite. At first this appears to be a contradiction. However, iron-rich smectites containing even higher amounts of iron have been reported.[15]

Table 3: Metal oxide content (wt%) determined by X-ray fluorescence spectroscopy and loss on ignition results

Metal oxide	Sample					
	V1	**U1**	**U2**	**U3**	**W1**	**W2**
SiO_2	42.4±0.4	52.8±0.4	47.6±0.4	53.9±0.4	46.8±0.4	78.2±0.4
TiO_2	2.00±0.03	1.75±0.03	2.11±0.03	1.73±0.03	2.19±0.03	0.80±0.03
Al_2O_3	19.6±0.3	14.3±0.3	18.2±0.3	15.1±0.3	20.4±0.3	7.76±0.3
Fe_2O_3	21.8±0.3	16.5±0.3	16.8±0.3	16.7±0.3	15.6±0.3	5.30±0.3
MnO	0.13±0.01	0.13±0.01	0.07±0.01	0.10±0.01	0.05±0.01	0.07±0.01
MgO	1.0±0.1	1.3±0.1	0.85±0.1	1.39±0.1	0.66±0.1	0.48±0.1
CaO	1.76±0.07	2.24±0.07	2.58±0.07	0.90±0.07	0.79±0.07	0.84±0.07
Na_2O	0.11±0.11	0.39±0.11	0.90±0.11	1.06±0.11	1.11±0.11	0.70±0.11
K_2O	0.06±0.06	0.24±0.06	0.10±0.06	0.20±0.06	0.11±0.06	0.76±0.06
P_2O_5	0.06±0.08	0.04±0.08	0.03±0.08	0.05±0.08	0.01±0.08	0.01±0.08
Cr_2O_3	0.05±0.01	0.05±0.01	0.04±0.01	0.05±0.01	0.03±0.01	0.02±0.01
V_2O_5	0.07±0.00	0.06±0.00	0.08±0.00	0.06±0.00	0.08±0.00	0.02±0.00
ZrO_2	0.02±0.00	0.04±0.00	0.06±0.01	0.06±0.01	0.05±0.01	0.06±0.01
Loss on ignition	12.06	11.01	11.92	9.63	12.10	4.94

The trace metals detected by XRF are listed in Table 4. Fortunately, all the samples are essentially free of the toxic elements As, Pb, Hg, Cd, Se and Sb. However, they do contain low levels of chromium and other heavy metals such as Cu, Zn and Ni. They do not contain detectable levels of the radioactive elements uranium and thorium. Strontium (Sr) (30–98 ppm), zirconium (Zr) (172–387 ppm) and yttrium (Y) (10–44 ppm) are present in substantial amounts. The stable form of Sr is not radioactive and it does not pose a significant human health risk at the levels detected. However, Zr and Y do pose a radioactive risk. Future studies should investigate whether the potential toxic elements are readily exchangeable or strongly bound to the mineral structures.

Additional data on the materials are presented in Table 5. Dispersion of sample V1 in water rendered it slightly acidic, but all the other samples produced a slightly basic pH. Clays used in cosmetics are usually fairly basic with the pH ranging from pH 7 to 10.5.[7] Literature reports indicate skin pH values ranging from pH 4 to 7, but 'natural' skin actually has a surface pH just below 5.[6,16] Apparently this pH is beneficial for the natural resident flora.[7] Application of clay-based cosmetics could make it too alkaline which could cause the skin to become too dry and sensitive, and may even result in eczema. So care is advised when using the present clays in skin applications for extended periods of time.

Table 4: Trace metal content determined by X-ray fluorescence spectroscopy

Element (ppm)	Sample					
	V1	**U1**	**U2**	**U3**	**W1**	**W2**
As	0	0	7	0	0	2
Cu	226	97	107	92	122	38
Ga	25	19	29	22	27	7
Mo	0	0	0	0	0	0
Nb	4	8	14	8	11	6
Ni	133	111	87	71	88	31
Pb	0	0	0	0	0	0
Rb	0	9	8	9	6	34
Sr	60	57	38	98	30	44
Th	0	0	0	0	0	0
U	0	0	0	0	0	0
W*	1	9	9	1	6	88
Y	31	29	44	34	41	10
Zn	71	50	41	49	43	29
Zr	172	279	387	350	351	317

**Semiquantitative estimates*

Table 5: Cation exchange capacity, estimate of upper bound for organic content and pH

	Sample					
	V1	**U1**	**U2**	**U3**	**W1**	**W2**
CEC (mEq/100 g)	30	31	37	37	35	19
TGA organics (wt%)	8.9	6.3	6.6	6.3	6.2	2.6
pH	6.1	7.9	9.1	8.0	8.7	8.6

CEC, cation exchange capacity; TGA, thermogravimetric analysis

The fairly significant loss on ignition values (9–12%) (see Table 3) suggest the presence of organic material. In some cases, when soil samples were heated in air, mass loss was observed in the TGA traces over the temperature range 200–500 °C (Figure 2). This mass loss often is associated with the oxidation of any organic substances present.[17,18] However, goethite decomposes into haematite at about 300 °C.[19] The mass loss resulting from the corresponding loss of water is 10.0 wt%. Taking this loss into account, an upper bound for the organic content of the materials was determined and these values are listed in Table 5. The apparent organic content varied from 2.6 wt% for sample W2 to 8.9 wt% for sample V1.

However, the absence of any bands near 3000 cm^{-1} in the FTIR spectra shown in Figure 3 suggests that very little, if any, organic material is actually present. The spectra can be fully explained by the presence of inorganic components in terms of previous observations.[20-23] The O-H asymmetric stretching occurs around 3700 cm^{-1}, probably as a result of the presence of the surface and internal OH groups of the Al-OH in the octahedral sheets. This band is lacking in the spectrum of sample W2, confirming the XRD result of the absence of kaolinite. The band at 3660 cm^{-1} is attributed to water OH bending vibrations[23], but there is only a very shallow dip in the spectrum of sample W2. The absence of a band located at 1230–1280 cm^{-1} implies the absence of amorphous silica. Si-O stretch vibrations are observed at 1000 cm^{-1}.[23] The bands near 920 cm^{-1} are typical for Si-O stretching and Al_2O-H deformations in the aluminium silicate, i.e. kaolinite.[20] The bands at 1100 cm^{-1} are assigned to Si-O and those at 1025 cm^{-1} to Si-O planar stretching. The bands around 775 cm^{-1} and 750 cm^{-1} can be attributed to the Al-O-Si inner surface vibration. The bands at 800 cm^{-1}, 775 cm^{-1} and 690 cm^{-1} in the spectrum of sample W2 are typical for quartz.[21]

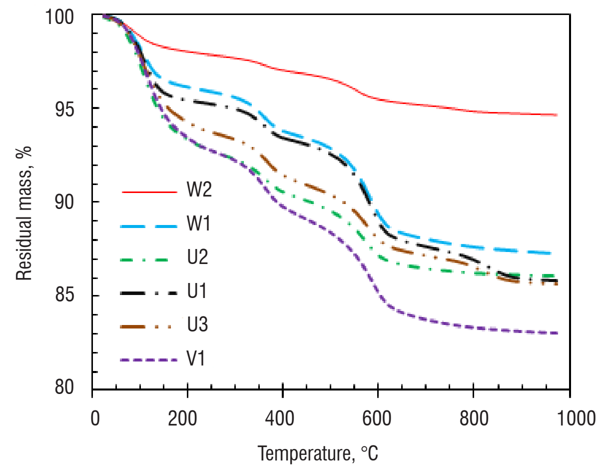


Figure 2: Thermogravimetric analysis mass loss curves for clay samples.

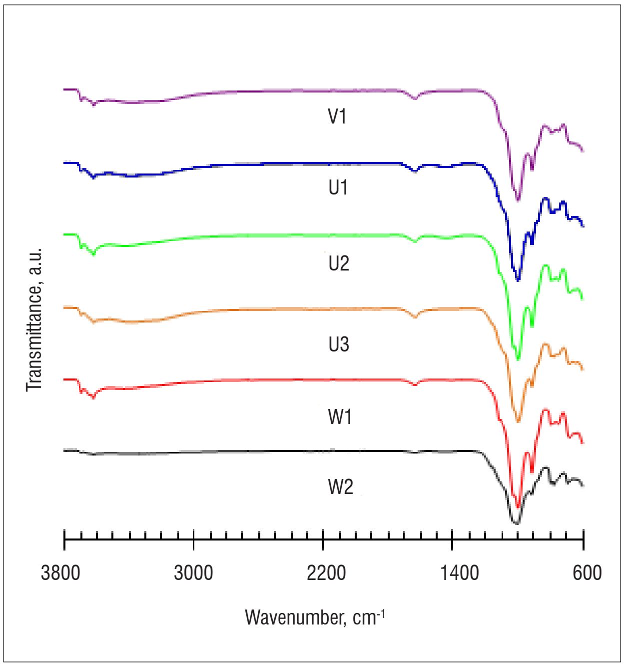
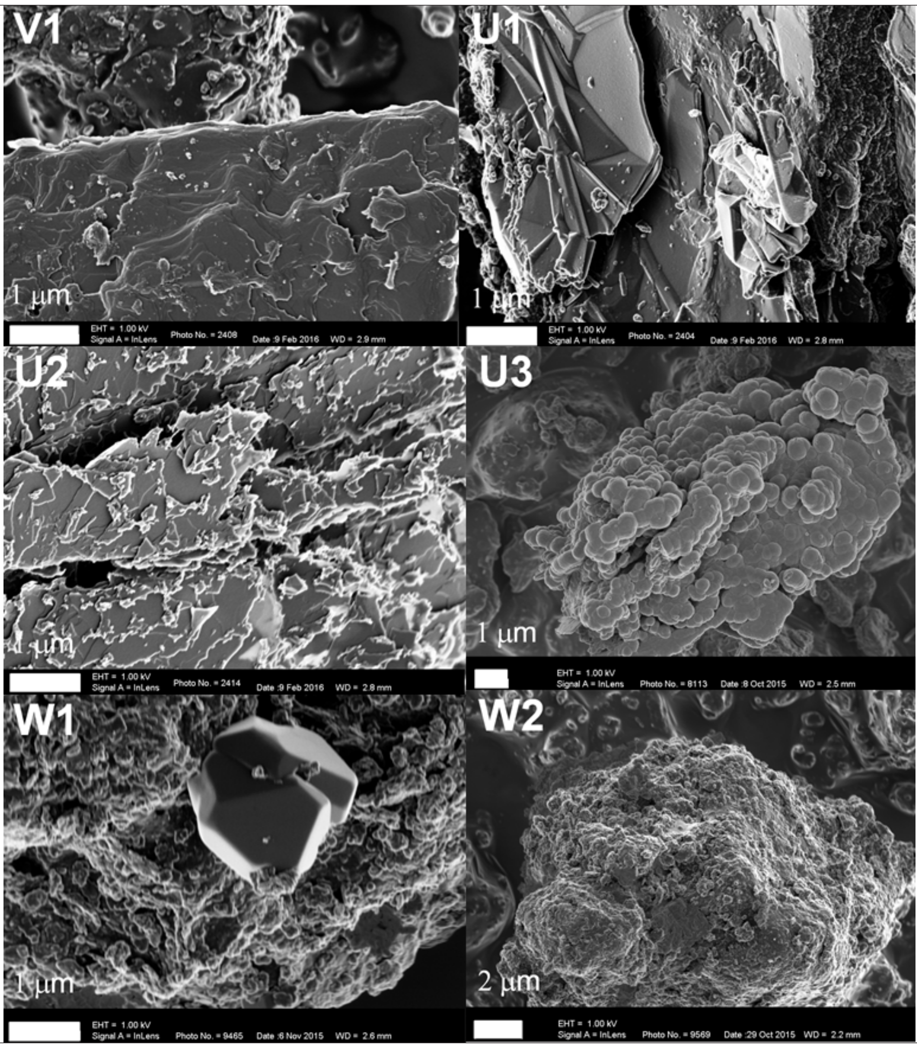


Figure 3: Attenuated total reflection Fourier transform infrared (ATR–FTIR) spectra of clay samples.



V1, U2 and U3: smectite platelets; U3: kaolin flakes; W1: a titanium dioxide crystal on top of quartz agglomerates; W2: quartz agglomerate

Figure 4: Scanning electron micrographs showing the morphology of powder particles.

The powder particles present in each individual clay sample exhibited various morphologies. The scanning electron micrographs in Figure 4 show only some selected particle morphologies found in each of the clay samples. However, in all cases, flake-shaped particles and highly agglomerated structures comprising small roundish sub-particles are present. The former reflect the presence of the phyllosilicates (kaolin, smectites, etc.), while the latter represent the other minerals present (silica, goethite, anatase, rutile, etc.).

Conclusion

We analysed clay samples sourced via traditional healers from Mtubatuba and Nongoma in KwaZulu-Natal and from Mbaleni in Limpopo, South Africa. These clays, known as *vumba* and *ubumba*, are supplied in a wet state and are reportedly used as topical cosmetics with health benefits. XRD analysis revealed a surprising variability in the mineralogical composition of the clays, even in samples taken from the same deposit. This implies that the actual application is robust with respect to clay composition. XRF analysis showed that the samples were free from radioactive elements and the toxic elements As, Pb, Hg, Cd, Se and Sb. However, they did contain low levels of chromium and other heavy metals, e.g. Cu, Zn, Ni. A disconcerting discovery is the high silica content, ranging from 11 to 85 wt%. Even though the intended purpose relates to topical applications, these levels may pose a certain health risk.

Acknowledgements

We thank Prof. T.S. Ramukumba and Ms V. Simelani-Ramela for help with the sourcing of the samples and Tshwane University of Technology for financial support including a sabbatical stay for R.M-M.

Authors' contributions

R.M-M. collected the samples, performed most of the experiments and managed the characterisation done by others, notably the XRD and XRF investigations. She also analysed the results and drafted the manuscript. W.W.F. supervised R.M-M., assisted in evaluating the experimental results and edited the manuscript. W.G. assisted with sample preparation (milling) and performed the XRD analyses.

References

1. Mackay A, Welz A. Engraved ochre from a Middle Stone Age context at Klein Kliphuis in the Western Cape of South Africa. J Archaeol Sci. 2008;35(6):1521–1532. http://dx.doi.org/10.1016/j.jas.2007.10.015

2. Henshilwood CS, D'Errico F, Van Niekerk KL, Coquinot Y, Jacobs Z, Lauritzen SE, et al. A 100,000-year-old ochre-processing workshop at Blombos Cave, South Africa. Science. 2011;334(6053):219–222. http://dx.doi.org/10.1126/science.1211535

3. Dayet L, Texier PJ, Daniel F, Porraz G. Ochre resources from the Middle Stone Age sequence of Diepkloof Rock Shelter, Western Cape, South Africa. J Archaeol Sci. 2013;40(9):3492–3505. http://dx.doi.org/10.1016/j.jas.2013.01.025

4. Wadley L. Cemented ash as a receptacle or work surface for ochre powder production at Sibudu, South Africa, 58,000 years ago. J Archaeol Sci. 2010;37(10):2397–2406. http://dx.doi.org/10.1016/j.jas.2010.04.012

5. Dlova NC, Nevondo FT, Mwangi EM, Summers B, Tsoka-Gwegweni J, Martincigh BS, et al. Chemical analysis and in vitro UV-protection characteristics of clays traditionally used for sun protection in South Africa. Photodermatol Photoimmunol Photomed. 2013;29(3):164–169. http://dx.doi.org/10.1111/phpp.12042

6. Mpako MP, Matike EM, Ekosse GI, Ngole VE. Ceremonial usage of clays for body painting according to traditional Xhosa culture. Indilinga. 2011;10(2):235–244.

7. López-Galindo A, Viseras C, Cerezo P. Compositional, technical and safety specifications of clays to be used as pharmaceutical and cosmetic products. Appl Clay Sci. 2007;36(1–3):51–63. http://dx.doi.org/10.1016/j.clay.2006.06.016

8. Matike DME, Ekosse GI, Ngole VM. Indigenous knowledge applied to the use of clays for cosmetic purposes in Africa: An overview. Indilinga. 2010;9(9):2138–2150.

9. Matike DME, Ekosse GIE, Ngole VM. Physico-chemical properties of clayey soils used traditionally for cosmetics in Eastern Cape, South Africa. Int J Phys Sci. 2011;6(33):7557–7566. http://dx.doi.org/10.5897/IJPS11.1249

10. Mpuchane SF, Ekosse GIE, Gashe BA, Morobe I, Coetzee SH. Microbiological characterisation of southern African medicinal and cosmetic clays. Int J Environ Health Res. 2010;20(1):27–41. http://dx.doi.org/10.1080/09603120903254025

11. Jumbam ND. Mineralogical and geochemical analyses of the healing elements in clayey soils from Isinuka traditional spa in Port St Johns, South Africa. Trans Roy Soc S Afr. 2014;68(1):25–31. http://dx.doi.org/10.1080/0035919X.2012.756075

12. Mattioli M, Giardini L, Roselli C, Desideri D. Mineralogical characterization of commercial clays used in cosmetics and possible risk for health. Appl Clay Sci. 2016;119:449–454. http://dx.doi.org/10.1016/j.clay.2015.10.023

13. Kahr G, Madsen FT. Determination of the cation exchange capacity and the surface area of bentonite, illite and kaolinite by methylene blue adsorption. Appl Clay Sci. 1995;9(5):327–336. http://dx.doi.org/10.1016/0169-1317(94)00028-O

14. Khiari I, Mefteh S, Sánchez-Espejo R, Cerezo P, Aguzzi C, López-Galindo A, et al. Study of traditional Tunisian medina clays used in therapeutic and cosmetic mud-packs. Appl Clay Sci. 2014;101:141–148. http://dx.doi.org/10.1016/j.clay.2014.07.029

15. Brigatti MF. Relationships between composition and structure in Fe-rich smectites. Clays Clay Miner. 1983;18(2):177–186. http://dx.doi.org/10.1180/claymin.1983.018.2.06

16. Lambers H, Piessens S, Bloem A, Pronk H, Finkel P. Natural skin surface pH is on average below 5, which is beneficial for its resident flora. Int J Cosmet Sci. 2006;28(5):359–370. http://dx.doi.org/10.1111/j.1467-2494.2006.00344.x

17. Kristl M, Muršec M, Šuštar V, Kristl J. Application of thermogravimetric analysis for the evaluation of organic and inorganic carbon contents in agricultural soils. J Therm Anal Calorim. 2016;123(3):2139–2147. http://dx.doi.org/10.1007/s10973-015-4844-1

18. Smidt E, Lechner P. Study on the degradation and stabilization of organic matter in waste by means of thermal analyses. Thermochim Acta. 2005;438(1–2):22–28. http://dx.doi.org/10.1016/j.tca.2005.08.013

19. Song S, Jia F, Peng C. Study on decomposition of goethite/siderite in thermal modification through XRD, SEM and TGA measurements. Surf Rev Lett. 2014;21(1), Art. #1450019, 6 pages. http://dx.doi.org/10.1142/S0218625X1450019X

20. Fadil-Djenabou S, Ndjigui P-D, Mbey JA. Mineralogical and physicochemical characterization of Ngaye alluvial clays (Northern Cameroon) and assessment of its suitability in ceramic production. J Asian Ceram Soc. 2015;3(1):50–58. http://dx.doi.org/10.1016/j.jascer.2014.10.008

21. Akyuz S, Akyuz T, Basaran S, Bolcal C, Gulec A. Analysis of ancient potteries using FT-IR, micro-Raman and EDXRF spectrometry. Vibrat Spec. 2008;48(2):276–280. http://dx.doi.org/10.1016/j.vibspec.2008.02.011

22. Russell J, Fraser A, Wilson M. Clay mineralogy: Spectroscopic and chemical determinative methods. London: Chapman Hall; 1994.

23. Boulingui JE, Nkoumbou C, Njoya D, Thomas F, Yvon J. Characterization of clays from Mezafe and Mengono (Ne-Libreville, Gabon) for potential uses in fired products. Appl Clay Sci. 2015;115:132–144. http://dx.doi.org/10.1016/j.clay.2015.07.029

Potential for identifying plant-based toxins on San hunter-gatherer arrowheads

AUTHORS:
Madelien Wooding[1]
Justin Bradfield[2,3]
Vinesh Maharaj[1]
Dwayne Koot[1]
Lyn Wadley[3]
Linda Prinsloo[4,5]
Marlize Lombard[2]

AFFILIATIONS:
[1]Department of Chemistry, University of Pretoria, Pretoria, South Africa
[2]Department of Anthropology and Development Studies, University of Johannesburg, Johannesburg, South Africa
[3]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa
[4]Department of Physics, University of Pretoria, Pretoria, South Africa
[5]Centre for Archaeological Science, School of Earth and Environmental Sciences, University of Wollongong, Sydney, Australia

CORRESPONDENCE TO:
Justin Bradfield

EMAIL:
Jbradfield8@gmail.com

DATES:
Received: 14 July 2016
Revised: 21 Oct. 2016
Accepted: 02 Nov. 2016

KEYWORDS:
San hunting poisons; southern Africa; liquid chromatography; accurate mass-mass spectrometry; archaeological analysis

HOW TO CITE:
Wooding M, Bradfield J, Maharaj V, Koot D, Wadley L, Prinsloo L, et al. Potential for identifying plant-based toxins on San hunter-gatherer arrowheads. S Afr J Sci. 2017;113(3/4), Art. #2016-0210, 10 pages. http://dx.doi.org/10.17159/sajs.2017/20160210

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa)

The antiquity of the use of hunting poisons has received much attention in recent years. In this paper we present the results of a pilot study designed to detect the presence of organic compounds, typically of less than 1200 Da, from poisonous plants that may have been used as hunting poisons in the past. We used ultra-performance liquid chromatography connected to a Synapt G2 high-resolution MS-QTOF mass spectrometer (UPLC-QTOF-MS) to provisionally identify plant-based toxins present in (1) extracts of fresh plant material, (2) a blind control recipe consisting of three plant ingredients and (3) a Hei||om arrow poison of unknown ingredients. Although not all expected toxic compounds were identified, those that were identified compared favourably with those reported in the literature and confirmed through databases, specifically the Dictionary of Natural Products and ChemSpider. MS/MS fragmentation patterns and accurate mass were used for tentative identification of compounds because archaeological residues usually contain insufficient material for unambiguous identification using nuclear magnetic resonance. We highlight the potential of this method for accurately identifying plant-based toxins present on archaeological artefacts and unique (albeit non-toxic) chemical markers that may allow one to infer the presence of toxic plant ingredients in arrow poisons. Any chemical study of archaeological material should consider the unique environmental degradative factors and be sensitive to the oxidative by-products of toxic compounds.

Significance:

- Methodology is presented for the identification of ancient plant-based arrow poisons.

Introduction

Bow hunting with poisoned arrows is well documented among southern African San hunter-gatherers.[1-3] Less well known is the great variety of toxic plants that were – or could have been – used for this purpose.[4-6] Most chemistry studies in which the ingredients of San poison arrows have been investigated focused on the more commonly known Chrysomelidae family of leaf beetle, which includes the *Diamphidia* and *Polyclada* genera.[7-10] Few studies thus far have been devoted specifically to plant poison ingredients,[11,12] and most of these studies are now several decades old. Now with the availability of new, more advanced, sensitive and reliable chemical detection techniques, we may be able to identify plant-based toxins present on archaeological artefacts.

Among the Kalahari San, large game is hunted with a bow and poisoned arrows[2,13] (Figure 1). This practice is widely considered to extend back at least 12 000 years,[14] but might be considerably older[15]. The identification of bow and arrow hunting systems can highlight aspects of technological complexity and past cognition.[16,17] More subtle innovations within hunting systems, such as the introduction of poisons, also have potential to inform on past cognitive frameworks and the time-depth of indigenous knowledge systems.[4] However, tracing such techno-behaviours through the Stone Age is not an easy task. The interpretation of bone artefacts as arrow components from 37 000 year-old levels at White Paintings Shelter[18], 43 000 year-old levels at Border Cave[15] and >61 000 years ago at Sibudu[19] pushes back in time the probable invention of bow and arrow hunting. This inference is corroborated by the functional interpretations of small quartz artefacts from Sibudu and Umhlatuzana dating to between 65 000 and 60 000 years ago.[20-23] The great antiquity of the use of certain toxic plants to poison arrows is purported from 24 000-year-old levels at Border Cave, KwaZulu-Natal, South Africa.[15] At this site, the remains of ricinoleic acid – an oxidative by-product of ricin – were discovered on a wooden stick, morphologically similar to 20th-century San arrow poison applicators.[15] Even earlier than this discovery, at approximately 77 000 years ago, people at Sibudu Cave (KwaZulu-Natal) constructed their bedding from plants with natural insecticidal and larvicidal properties[24], implying a practical understanding of the biochemical properties of certain plants.

In this paper we build on the growing interest in ancient poison chemical characterisation[25] and present the results of a pilot study designed to detect the presence of small organic compounds, typically of less than 1200 Da (mass in Dalton units), from poisonous plants. An ultra-performance liquid chromatography system coupled to a Synapt G2 quadrupole time-of-flight mass spectrometer (UPLC-QTOF-MS) was used provisionally to identify known toxic compounds through comparison of their accurate masses to those recorded in databases (such as ChemSpider and Dictionary of Natural Products) for 11 of the most commonly occurring toxic plants reported in the southern African literature. Poison from a 100-year-old bone arrow from northern Namibia as well as a blind control poison recipe were subsequently analysed as a proof of concept using the same technique. We highlight the potential of this method for accurately identifying plant-based toxins present on archaeological artefacts.

Background

A diverse group of organic compounds known as secondary metabolites is produced by plants. Secondary metabolites defend plants against a variety of pathogens and herbivores[26] and are of great importance for medicinal drugs, industrial materials and poisons[27].

Photo: ©Lyn Wadley

Figure 1: Ju/'hoan hunter in a Nyae Nyae village mixing poison in a hollow bone. The poison is applied to the new arrowheads lying in front of him. *Swartzia madagascariensis* pods are next to the glue stick/poison applicator.

Alkaloids, terpenoids and cardiotonic glycosides, the main secondary metabolite compounds responsible for the toxicity of arrow poisons[26,28], are regularly used in small doses for their medicinal qualities in modern pharmaceuticals[29,30] and have a long history of ethnopharmacological use[12]. Although currently much is known about the active toxins in some southern African flora species[12,26], many species remain insufficiently studied[4]. In some cases, studies have focused only on certain parts of plants such as the leaves or fruit rather than the sap and roots. Chemical compounds are distributed differently throughout the plant[29] – a phenomenon of which the San were aware, as they would only use the toxic parts such as the sap of succulent plants[4,12].

San hunters sometimes included additives with their arrow poison. A variety of reasons for this practice has been offered, for example, to increase the viscosity of the mixture to aid adhesion and to enhance the efficacy of the poison.[30,31] The additives result in complex recipes containing multiple organic components derived from different sources,[4] which makes trying to identify specific toxins in a sample of ancient poison challenging. Not only will one expect to see the by-products of the oxidative breakdown of the toxic compounds, but identification will be complicated by the combination of numerous compounds.[25]

Archaeological samples can be divided into inorganic materials and organic materials. Inorganic materials include stone tools, pottery and metal artefacts. Organic materials can include plant and animal remains as well as their deteriorative and biotransformative products. Inorganic materials are generally better preserved than organic materials. Preservation of organic materials such as plant toxins is significantly affected by environmental factors, such as soil pH, temperature, oxygen, moisture exposure and substrate, especially if the sample has been exposed to these factors over extended periods of time.[32] The rate of decomposition of organic compounds is variable depending on the class and structural type. For example, sugars, starches and simple proteins may decompose at a faster rate than lignins and phenolic compounds. As such, phytochemical residues present on archaeological artefacts may not perfectly resemble a freshly extracted phytochemical profile of the parent compound. Any chemical study of archaeological material should therefore consider the oxidative by-products of toxic compounds, which further complicate the analyses as a result of the complex matrix.

Materials and methods

Sample preparation

Eleven South African plants

Eleven plants were selected for this study based on previously reported studies that show them to contain toxic compounds of known chemical structures, and therefore with known molecular formulae (Table 1). Fresh plants were collected mainly from the Walter Sisulu National Botanical Gardens and the parts traditionally used for poison preparation by the San[4] were air dried and used for the analysis.

Blind test on the control sample

The second part of our study involved a blind test. A poison mixture was prepared in the laboratory, mimicking a known San poison recipe. This recipe, consisting of three poisonous plant ingredients, was prepared in accordance with historical San practices.[4,11] The ingredients of this recipe were known to only one author (JB), and the sample was prepared as follows:

1. The stem and leaves of *Acokanthera oppositifolia* were boiled in water for 8 h until a yellow viscid fluid was obtained.
2. Once cooled, the latex of *Euphorbia tirucalli* was added.

3. To this mixture was added the juice of *Adenium multiflorum*, which was obtained by heating a branch of *A. multiflorum* over a fire until the juices oozed out of the cut end.
4. The ingredients were collected in a glass Petri dish and mixed.
5. The resulting beige liquid was applied to replica stone and bone artefacts and left in the sun to dry for 24 h.
6. Once dry, the consistency was that of a hard resin.

These 'tools' were chemically analysed approximately 2 weeks after preparation.

Analysis of poison on a Hei||om arrow

The third part of our study was the examination of a poison sample of unknown chemistry taken from a Hei||om arrow, purportedly collected by Dr Louis Fourie[33,34] in Namibia[35,36]. This arrow, made available by Museum Africa to JB, is unaccessioned, but was found with the Fourie collection and is identical to the rest of the Hei||om bone-tipped arrows. It is considered to be of equivalent age (± 90 years), and most likely part of the same collection.

Extraction, isolation and analysis

Eleven South African plants

Each plant sample was extracted, isolated and analysed in the same way. Each plant was air dried and ground to a fine powder. An amount of 1 g of the dried powdered plant material was extracted by stirring for 1 h with 10 mL of dichloromethane:methanol (DCM:MeOH, 1:1). The extraction solvents were of analytical grade and purchased from Fluka. The extraction was repeated twice and filtered through Whatman filter paper no. 1 before pooling of the solvent extracts. The extracts were dried under reduced pressure below 40 °C using a rotary evaporator. Samples were stored dry in a temperature-controlled room at 23 °C prior to analysis.

Blind control sample

The replicated stone and bone artefacts held only small quantities of poison. Therefore 1 mL of DCM:MeOH (1:1) was added directly to the containers holding the poison. Extraction was done by ultrasonication of the mixture for 30 min. The extract was filtered and solvents were removed in vacuo. Samples were stored in a temperature-controlled room (23 °C) prior to analysis.

Table 1: Fresh plant samples listing the reported toxins and their accurate mass

Sample #	Plant name	Plant parts used	Type of toxin	Toxic compound
MW-1-4A	*Acokanthera oppositifolia* (Bushman's poison)	Leaves and stem	Cardiac glycoside	Ouabain ($C_{29}H_{44}O_{12}$) Acovenoside A ($C_{30}H_{46}O_9$)
MW-1-2B	*Adenium multiflorum* (Impala lily)	Stem	Cardiac glycoside	Obebioside B ($C_{38}H_{58}O_{15}$) Hongheloside B ($C_{36}H_{56}O_{14}$) Tetraphyllin B ($C_{12}H_{17}NO_7$)
MW-1-3C	*Aloe gariepensis*	Sap		γ-Coniceine ($C_8H_{15}N$) Conhydrine ($C_8H_{17}NO$)
MW-1-3F	***Aloe globuligemma***	Sap	Piperidine	Coniine ($C_8H_{17}N$) Conhydrine ($C_8H_{17}NO$)
MW-1-3A	*Ammocharis coranica* (Karoo lily)	Bulb	Isoquinoline alkaloid	Lycorine ($C_{16}H_{17}NO_4$) Caranine ($C_{16}H_{17}NO_3$) Crinamine ($C_{17}H_{19}NO_4$) Acetylcaranine ($C_{18}H_{19}NO_4$)
MW-1-2A	*Boophane disticha* (Poison bulb)	Bulb	Isoquinoline alkaloid	Haemanthamine/Crinamine ($C_{17}H_{19}NO_4$) Lycorine ($C_{16}H_{17}NO_4$) Buphanine ($C_{18}H_{21}NO_4$) Crinamidine ($C_{17}H_{19}NO_5$) Distichamine ($C_{18}H_{19}NO_5$)
MW-1-55A MW-1-77	*Euphorbia tirucalli* (Pencil plant)	Latex Leaves	Diterpenoid	Phorbol ($C_{20}H_{28}O_6$) Diterpene ($C_{20}H_{32}$)
MW-1-3E MW-1-55B	*Euphorbia ingens* (Candelabra tree)	Leaves Latex	Diterpenoid	Ingenol ($C_{20}H_{28}O_5$)
MW-1-3D	*Euphorbia virosa* (Poison tree)	Latex	Diterpenoid	Diterpene ($C_{20}H_{32}$)
MW-1-3B	***Strophanthus speciosus*** (Poison rope)	Seeds	Cardiac glycoside	Ouabain ($C_{29}H_{44}O_{12}$) Christyoside ($C_{30}H_{44}O_9$)
MW-1-4B	*Strychnos madagascariensis* (Black monkey orange)	Unripe seeds	Indole alkaloid	C-toxiferine I ($C_{40}H_{46}N_4O_2$)

Note: Species listed in bold font are considered to be lethally toxic in relatively small quantities. A complete reference list for the toxic compounds identified in these plants can be found elsewhere.[4,12]

Poison on a Hei||om arrow

Approximately 2 mg of material was scraped off the 90-year-old Hei||om arrow and added to 1 mL DCM:MeOH (1:1). The sample was then treated as for the control sample.

Ultra-performance liquid chromatography QTOF mass spectrometry

Compound separation and detection were performed using a Waters UPLC hyphenated with a Waters Synapt G2 QTOF instrument. The DCM:MeOH dried extracts were reconstituted, first in 100% acetonitrile followed by water (0.1% formic acid) such that the final concentration was ~1 mg/mL of total crude extract. MS-grade acetonitrile was purchased from Romil. Water with 0.1% formic acid was purchased from Sigma Aldrich. The extracts were pooled and centrifuged at 10 000 g for 10 min to remove particulates. Prior to analyses, the instrument was calibrated over a mass range of 50–1200 Da using a sodium formate solution, typically to an absolute mass accuracy of <0.5 mDa using the Intellistart functionality of the software. The instrument was centrally operated and controlled with MassLynx v4.1 software for data acquisition. A form of data independent analysis termed MS^E was used to acquire both low energy (precursor ions) and high energy (product ions) utilising a collision energy ramp from 10 V to 40 V over a scan time of 0.3 s. An internal control (the lockspray), namely leucine encephalin, was directly infused into the source through a secondary orthogonal electrospray ionisation probe allowing intermittent sampling (every 10 s). The lockspray was used to compensate for instrument drift, thus ensuring good mass accuracy throughout the duration of the runs. Exactly 5 µL of the reconstituted extracts was injected into the UPLC-MS system. All the samples were run in both positive and negative ionisation modes (Table 2).

Table 2: Parameters of the ultra-performance liquid chromatography quadrupole time-of-flight mass spectrometer (UPLC-QTOF-MS) system

Liquid chromatography system	Acquity®
Detector	Waters Synapt G2QTOF
Calibration mass range	50–1200 *m/z* using sodium formate clusters and Intellistart functionality
Capillary	2.8 kV
Ionisation mode	Electrospray ionisation
Source temperature	100 °C
Sampling cone	15 V
Extraction cone	4 V
Desolvation temperature	200 °C
Cone gas flow	100 L/h
Desolvation gas flow	500 L/h
Column	Waters C18 BEH, 1.7 µm particle size
Elution scheme	30-min gradient elution scheme from 98% H_2O (0.1% formic acid) to 100% acetonitrile (0.1% formic acid)
Resolution	~20 000 FWHM
Absolute mass error	<0.5 mDa

BEH, Ethylene Bridged Hybrid

Separation was completed using a reverse phase step gradient elution scheme from 97% H_2O (0.1% formic acid) to 100% acetonitrile (0.1% formic acid). The column temperature was kept constant at 40 °C and the flow rate was set at 0.4 mL/min for the entire run, giving a total run time of 20 min. A Waters UPLC® C_{18} Ethylene Bridged Hybrid 1.7 µm particle size (2.1 mm ID x 100 mm length) column was used. Extracted ion chromatograms of the monoisotopic masses for the reported toxic compounds (see Table 1) were obtained from the base peak ion chromatograms to determine the presence of the target compounds in the particular plant sample. The chromatograms indicated the pseudo-molecular ion peak because electrospray ionisation is typically achieved through the addition or removal of one or more protons, i.e. $[M+H]^+$. The mass of common adducts considered included sodium and potassium in positive mode as well as loss of water, methoxy and acetyl groups. Mass accuracy of precursor ions was used to generate elemental formulae which could then be searched and compared against those in the literature and databases. The acquired isotopic distribution patterns were compared to the proposed elemental formula as further confirmation. In addition, product ion spectra (MS/MS fragments) were obtained for the target compounds from the various plant extracts and blind test samples to further supplement matching of retention time and precursor masses giving a high degree of confidence. Fragmentation patterns of compounds tentatively identified from accurate mass were generated using ChemDraw version 8.0. The fragmentation was used to confirm the MS/MS data.

As a first comparative step, the base peak ion chromatograms of all the samples were overlaid, which allowed visual identification of any obvious common constituents. MarkerLynx and ChromaLynx version 4.1 software was then used to select compound peaks (molecular features possessing a unique retention time–accurate mass pair) from the data matrix through spectral deconvolution algorithms, and to compare similarities and differences among all the samples. The 11 plant extracts and the two poison sample extracts were run in duplicate and only peaks observed in both runs were considered. Method blanks for both the plant extracts and poison samples were run in duplicate for background subtraction purposes.

Results

Identification of toxins through UPLC-QTOF mass spectrometry

Known toxins reported in 6 of the 11 plants were tentatively identified based on their accurate mass and MS/MS fragmentations (product ions). However, the occurrence of isomers – structural- and stereoisomers (compounds with the same molecular formula) – makes it difficult to unambiguously identify compounds based on accurate mass and MS/MS fragmentation. The positive and negative ion chromatograms were obtained for all the plant samples. A selective ion search (extracted ion chromatogram) was done using the monoisotopic mass for the selected toxic compound and possible adducts, as listed in Table 3.

To demonstrate the accuracy and speed in identifying known compounds in complex matrices such as plant extracts, the UPLC positive mode base peak ion chromatogram of an extract of *Boophane disticha* is shown in Figure 2. The extracted ion chromatograms displaying the compound peaks for distichamine, haemanthamine/crinamine and buphanidrine (and their structures and isotopic distributions) are shown in Figure 3.

Analysis of archaeological samples and control

Chromatographic overlays of the arrow poison extract and the control recipe extracts were visually compared with all the plant extracts. However, trying to determine areas of overlap proved to be highly complex. The positive and negative base peak ion chromatograms for the extracts of the blind control recipe and arrow poison extracts are shown in Figure 4a and 4b, and indicate the complexity of these samples.

Table 3: Tentative identification of the pseudomolecular ions and adducts of the targeted toxic compounds and compounds not found

Plant	Toxic compound and molecular formula	Acquired [M+H]+	Theoretical [M+H]+	Calculated accurate mass (Da)	Mass error (ppm)
Acokanthera oppositifolia	Aglycone of acovenoside A ($C_{23}H_{34}O_5$)	391.2459	391.2479	390.2406	5.1
Adenium multiflorum	Aglycone of obebioside B ($C_{25}H_{36}O_6$)	433.2589	433.2584	432.2511	-1.2
	Aglycone of hongheloside B ($C_{23}H_{34}O_5$)	391.2459	391.2479	390.2406	5.1
Aloe gariepensis	γ-Coniceine ($C_8H_{15}N$)	126.1293	126.1277	125.1204	-12.7
Aloe globuligemma	Coniine ($C_8H_{17}N$)	128.1440	128.1434	127.1361	-4.7
Ammocharis coranica	Lycorine ($C_{16}H_{17}NO_4$)	288.1236	288.1230	287.1157	-2.1
	Caranine ($C_{16}H_{17}NO_3$)	272.1281	272.1281	271.1208	0.0
	Crinamine ($C_{17}H_{19}NO_4$)	302.1397	302.1387	301.1314	-3.3
	Acetylcaranine ($C_{18}H_{19}NO_4$)	314.1395	314.1387	313.1314	-2.5
Boophane disticha	Distichamine ($C_{18}H_{19}NO_5$)	330.1343	330.1336	329.1263	-2.1
	Haemanthamine/ Crinamine ($C_{17}H_{19}NO_4$)	302.1391	302.1387	301.1314	-1.3
	Lycorine ($C_{16}H_{17}NO_4$)	288.1233	288.1230	287.1157	-1.0
	Buphanine/ Buphanidrine ($C_{18}H_{21}NO_4$)	316.1544	316.1543	315.1470	-0.3
	Crinamidine ($C_{17}H_{19}NO_5$)	318.1349	318.1336	317.1263	-4.1

Results obtained through UPLC-MS and extracted ion chromatograms.

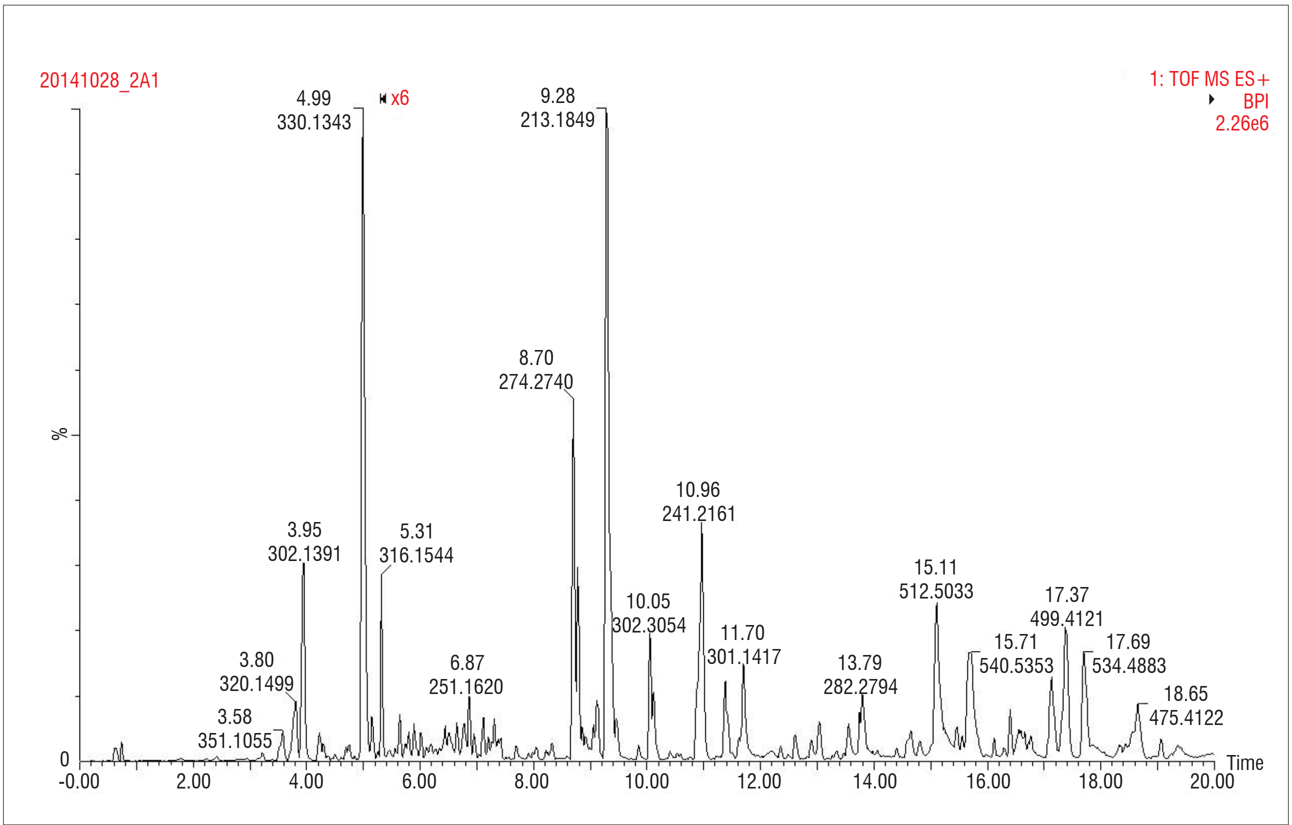


Figure 2: Positive mode base peak ion chromatogram of *Boophane disticha*.

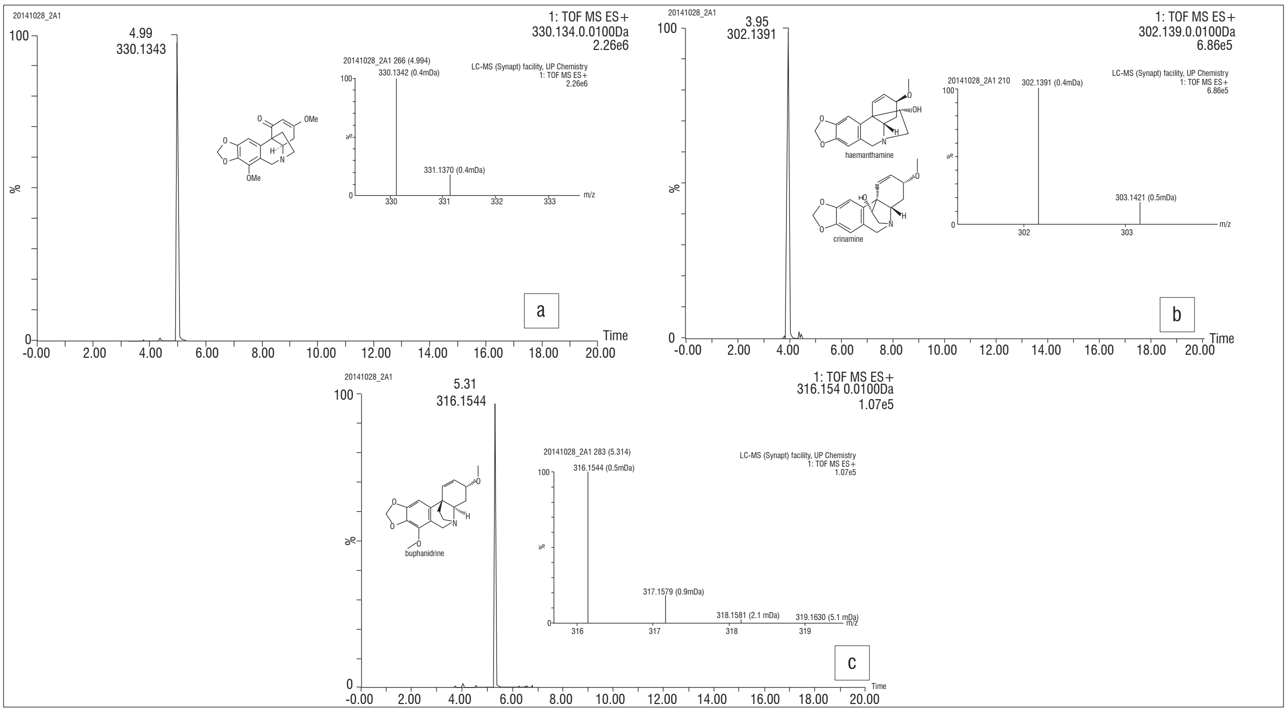


Figure 3: Extracted ion chromatogram from *Boophane disticha* showing (a) the extracted peak chromatogram of 330.1343 corresponding to the accurate mass of distachamine; (b) the extracted peak chromatogram of 302.1391 corresponding to the accurate mass of stereoisomers haemanthamine and/or crinamine and (c) the extracted peak chromatogram of 316.1544 corresponding to the accurate mass of buphanidrine. The structures and acquired isotopic distribution are shown in the insets.

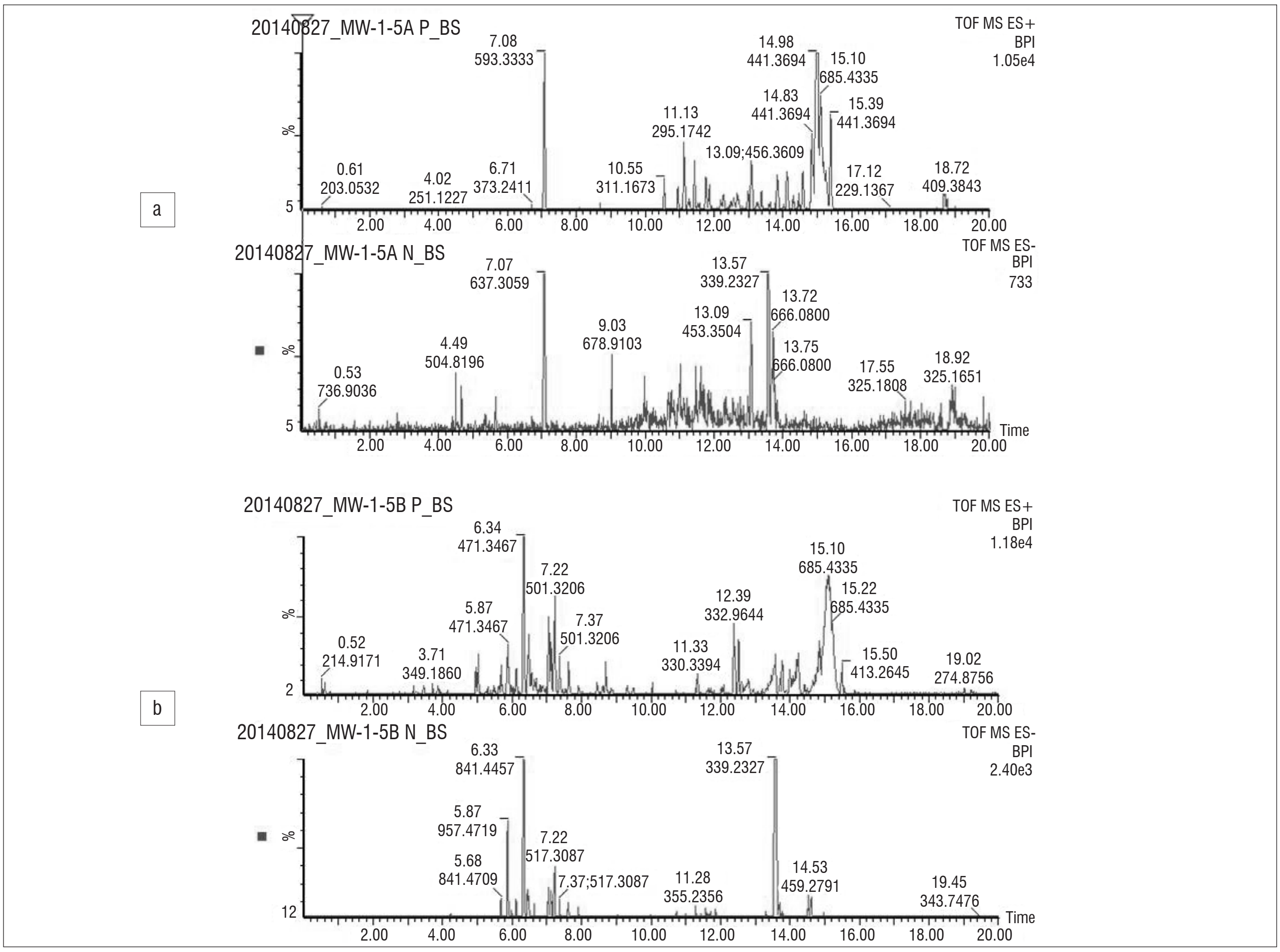


Figure 4: Positive (top) and negative (bottom) base peak ion chromatograms of (a) the control recipe and (b) the Hei||om arrow poison sample.

MarkerLynx software was used to identify molecular features (retention time–accurate mass pairs) between all samples and a total of 13 049 unique features was identified. The features were compared with ChromaLynx software to identify possible markers (molecular features common amongst a plant extract and a blind sample). Markers with a retention time of 7.04 min and *m/z* 1185.6515 from the extract of *A. multiflorum* and retention time of 14.49 min and *m/z* 967.6968 from the extract of *E. tirucalli* were both present in the blind control recipe (Figure 5). A marker detected using the software was present in both the blind control sample and *A. oppositifolia*. This compound, however, was not exclusive to these extracts because it was also detected in some of the other plant samples.

The comparison of the chromatographic overlays of the Hei||om arrow poison with all the plant extracts indicated a marker with a pseudomolecular ion, *m/z* 332.9644, common to the arrow poison extract and the extract of *Strychnos madagascariensis* (Figure 6). However, ChromaLynx software analysis revealed the presence of trace amounts of this ion in *Strophanthus speciosus, Euphorbia virosa, A. oppositifolia* and the control recipe, but not in the sample blank, implying that the Hei||om poison is a compound of plant origin. Although MS/MS fragmentation was obtained, no conclusive identification of the compound could be made based on the fragmentation patterns. The other compounds that appeared promising as possible markers through visual examination of the chromatograms between the arrow poison and *S. madagascariensis* were eliminated by the MarkerLynx software analysis. Only the compound with *m/z* 332.9644 and retention time 12.39 min was identified as a common feature in both the Hei||om poison and *S. madagascariensis*. The fragmentation of this molecular ion was identical between the arrow poison and the *S. madagascariensis* plant sample (Figure 7), providing further evidence of their similarity.

Discussion and conclusions

Here we report the results of chemical analyses conducted on three sets of material. Firstly, we used extracts from 11 modern plants supplied by the South African National Biodiversity Institute and collected from the Walter Sisulu National Botanical Gardens. Secondly, we created a toxic mixture from three plant extracts and used this mixture as a blind test control. Finally, we analysed the poison on an ethnographic artefact, a Hei||om arrowhead.

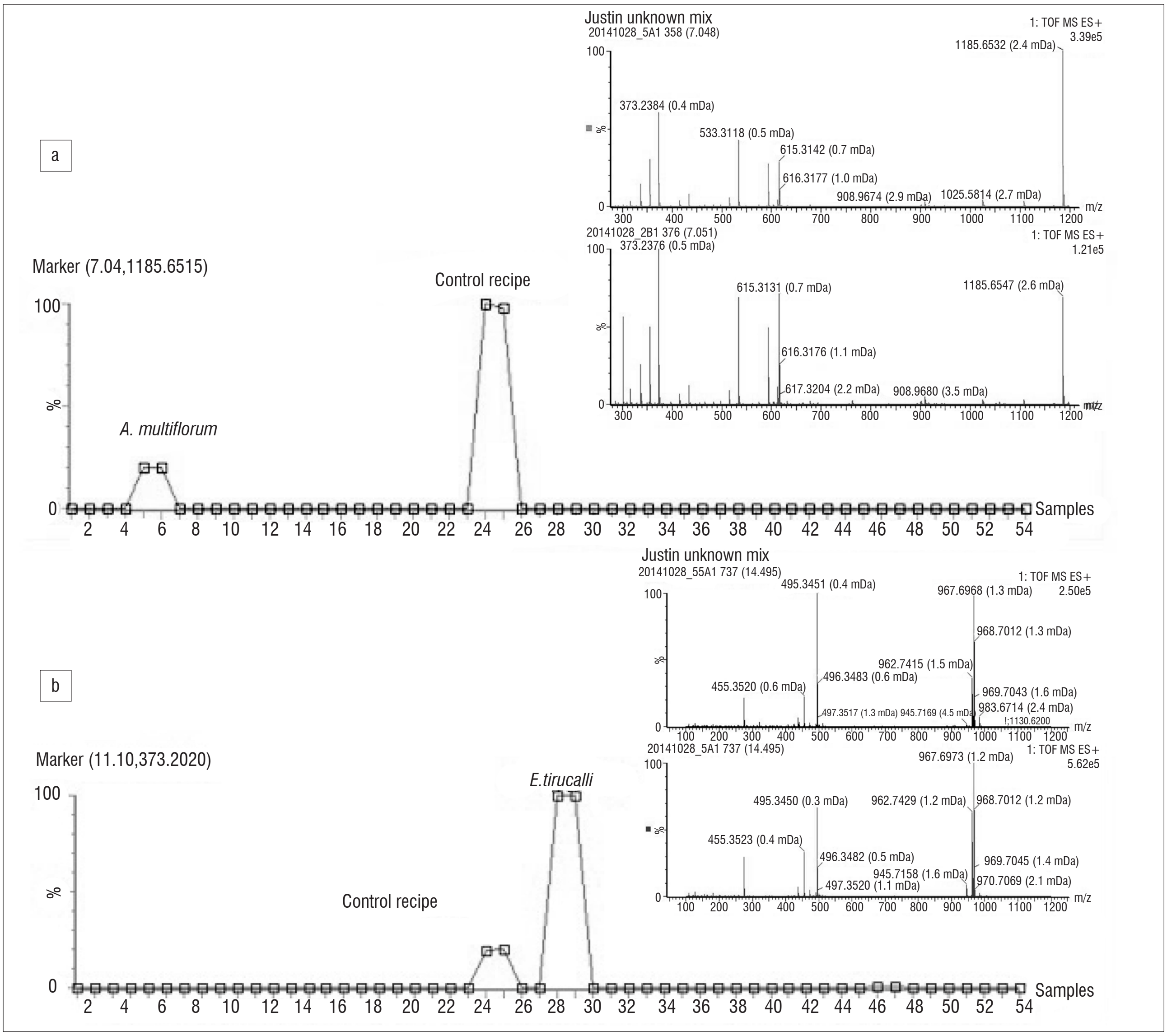


Figure 5: MarkerLynx results showing (a) a marker at 7.04 min and *m/z* 1185.6515 between *Adenium multiflorum* (sample numbers 5 and 6) and the control recipe (sample numbers 24 and 25) and (b) a marker at 14.495 min and *m/z* 967.6968 between *Euphorbia tirucalli* (sample numbers 28 and 29) and the control recipe (sample numbers 24 and 25). The mass spectra (top right) confirm the presence of a common compound in both extracts.

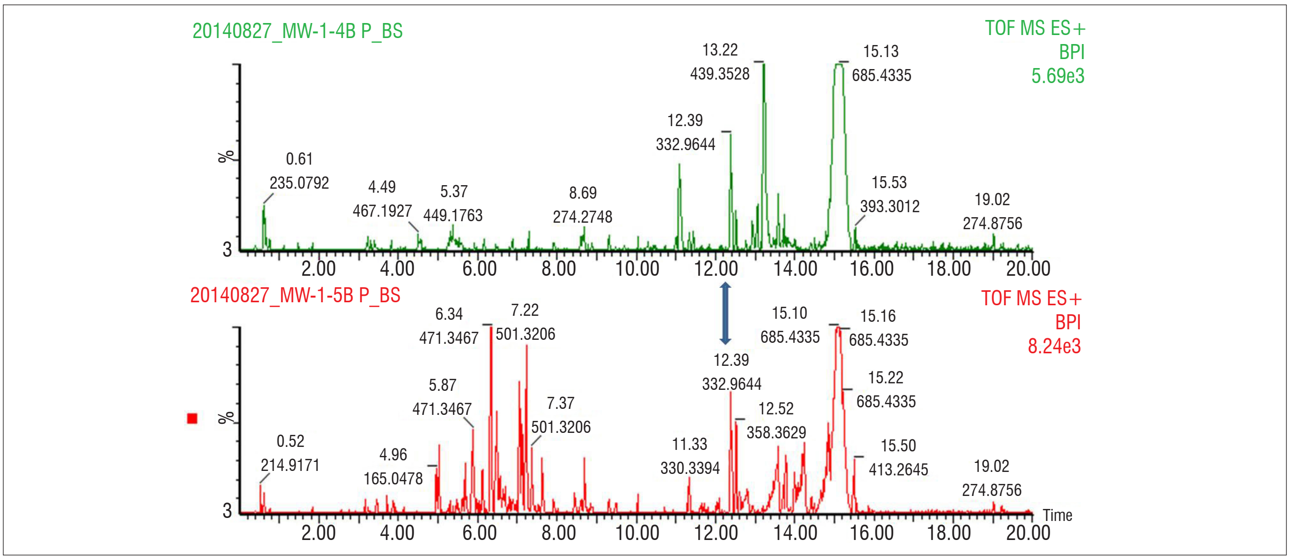


Figure 6: Positive base peak ion chromatograms of *Strychnos madagascariensis* (top) and the Hei||om arrow poison (bottom), visually indicating the presence of the common compound with pseudomolecular ion *m/z* 332. The molecular ions at retention times 15.10 min, *m/z* 685.4335, and at 19.02 min, *m/z* 274.8756, are a result of impurities or are nonspecific to the samples.

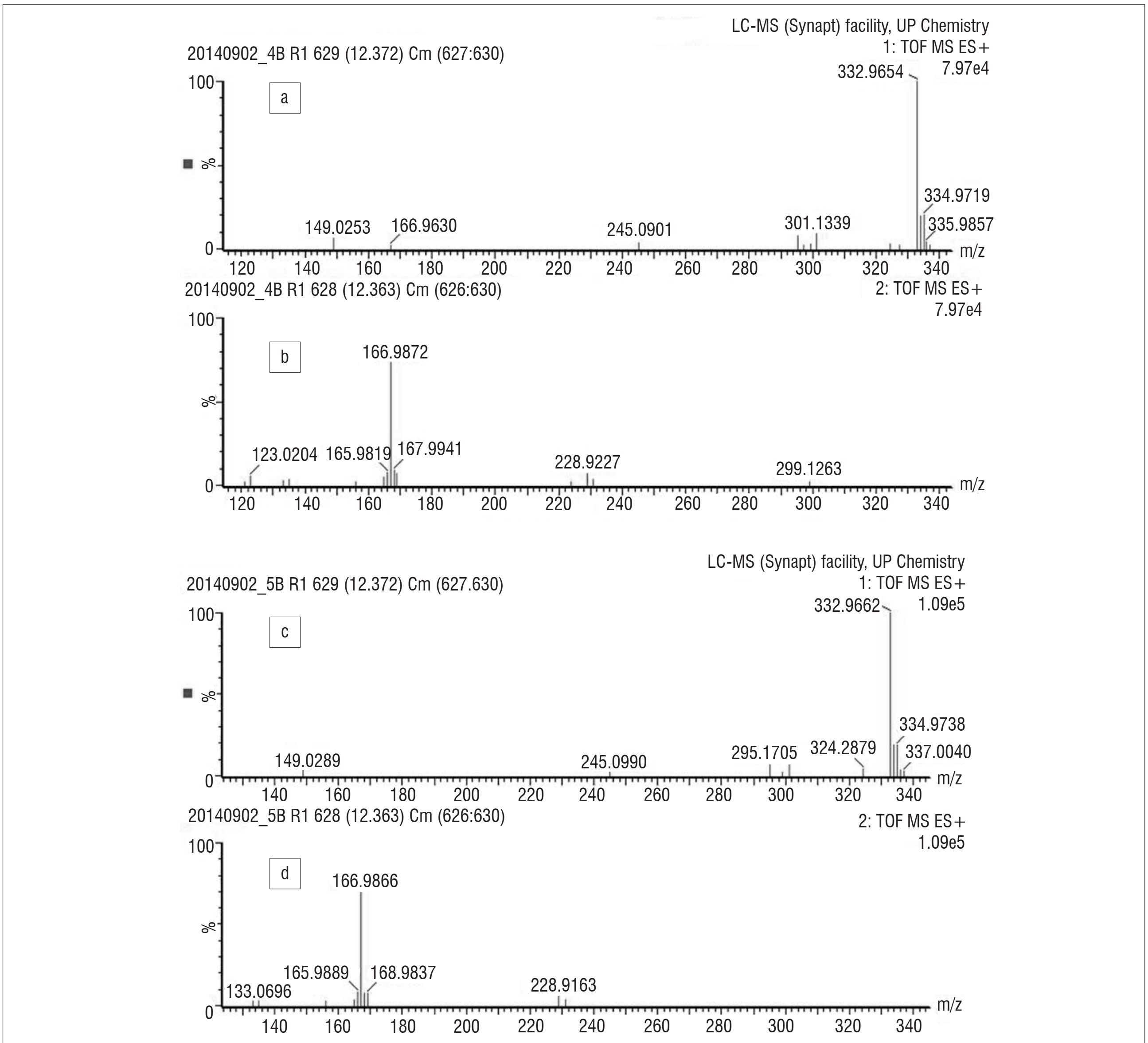


Figure 7: Mass-mass spectrometry fragments for *m/z* 332.9647 for (a and b) the Hei||om arrow poison and (c and d) *Strychnos madagascariensis* extract.

The results show that our method can be used to tentatively identify toxins based on comparative overlays with fresh plant material, but that no unambiguous associations can be made at this stage. Furthermore, not all expected toxins will be detected, for which there may be several reasons (which are discussed below). The identification of toxic compounds in plants has been typically conducted through the purification and isolation of organic compounds followed by detailed nuclear magnetic resonance analysis to elucidate their structures. This procedure requires large quantities of plant sample and is often a tedious and time-consuming process. This method cannot be used for determining the toxic plant compounds used in arrow poisons because of their small quantities. In order to facilitate the process using limited quantities of arrow poison samples, and to confirm the presence of reported toxins, we have used accurate MS/MS spectrometry. Our aim was to establish whether this technique can be applied to determine the presence of the toxic compounds in both known toxic plants and arrow poisons, which are sometimes mixtures of ingredients from several plants. The UPLC-QTOF-MS technique acquired mass spectral data of the reported toxic compounds in extracts of 11 toxic plants in both the positive and negative mode electrospray ionisation to ensure that most of the compounds were ionised and could be detected. We identified 16 of the 28 reported toxins in 6 of the 11 plants analysed. We did this by comparing the acquired accurate mass with the theoretical accurate mass based on the molecular formula. Further confirmation of the structure was achieved through the MassLynx software which also generated the molecular formula. The toxic compounds identified compared favourably with those in the databases Dictionary of Natural Products and ChemSpider, which report the accurate masses.

All identifications of the toxic compounds were made with high accuracy; they cannot, however, be unambiguously assigned. For unambiguous assignment, full MS/MS fragmentation analysis must be done and then compared to pure standard compounds and/or nuclear magnetic resonance analysis of the pure isolated compounds. Haemanthamine and crinamine, two previously published compounds from *B. disticha*, were identified in the plant extracts. We were unable to determine whether both, or either, were present as they are isomeric forms. The similar assignment problem applies to buphanine and buphandrine present in this plant extract. Two other toxic compounds, lycorine and crinamidine, were also identified in our *B. disticha* extract. Two of three toxins reported in published databases, namely the aglycones of obebioside B and hongheloside B, were identified in our *A. multiflorum* extract. All four of the previously published toxins (lycorine, caranine, crinamine and acetylcaranine) were identified in our extract of *Ammocharis coranica*. However, the two previously published toxins, ouabain and christyoside, could not be identified in our *S. speciosus* extract; γ-coniceine and coniine were identified from *Aloe gariepensis* and *Aloe globuligemma*, respectively, whereas conhydrine, common to both the *Aloe* species, was not identified. None of the toxic compounds previously published was identified in any of the *Euphorbia* species we analysed or in our sample of *S. madagascariensis.* Both ouabain and acovenoside A were identified in the highly toxic plant *A. oppositifolia.* The non-identification of toxic compounds in some of our plant extracts may be attributed to the geographical area in which the plants were collected. It is known that plant species display a variation in their secondary metabolites as a consequence of the environmental conditions and geographical location in which they are grown and the season in which they are harvested.[37] In addition, the generic extraction methodology and the use of MS friendly aqueous solvents (and additives) may have excluded them according to their solubility, or inhibited their preferred ionisation in negative mode.

To assess the feasibility of the method, we applied the UPLC-QTOF-MS approach to the detection of plant-based poisonous compounds in a control sample in which three plant extracts were mixed. Two of the three plants used in the control sample poison were correctly identified as *A. multiflorum* and *E. tirucalli*, based on unique markers, namely *m/z* 1185.6532 and 967.6968. Indiscriminate markers that are present in several plants were found in the control sample as well as in the extract of *A. oppositifolia*. This finding is not unusual as the same chemical compounds can occur in several different plant species.

The UPLC-QTOF-MS results of the Hei||om arrow poison were visually compared to chromatograms of the plants analysed with the aid of chromatographic overlays followed by analysis with MarkerLynx and ChromaLynx software. A marker with a pseudomolecular ion *m/z* 332.9650 was common to the Hei||om arrow extract and the extract of *S. madagascariensis*. ChromaLynx software analysis, however, revealed the presence of trace amounts of this ion in *S. speciosus, E. virosa, A. oppositifolia* and the control recipe. Although *S. madagascariensis* is not present in the region from which the arrow is thought to have come, it may be that the poisonous ingredient or the arrow itself was originally obtained through exchange from farther afield. Ideally, other species of *Strychnos*, such as the *S. spinosa*, should be collected and similarly analysed. It is encouraging that we were able to get chemical signatures from small samples and that there are clearly identifiable plant-based toxins detectable in poisons created more than 90 years ago. This bodes well for future work on ethnographic and archaeological material, despite the various challenges. These challenges include the limited quantities of arrow poison samples likely to be available for analysis from archaeologically recovered artefacts, and harsh post-depositional conditions that may have caused degradation of organic residues. Several other factors such as the variation of secondary metabolites between wild and cultivated plants, chemical variation between species, the production of certain classes of compounds being restricted to a specific plant part (e.g. roots, leaves), and the solvents used for the extraction of toxic compounds, all play an influential role and have to be taken into consideration.

Acknowledgements

Most of the plant material was supplied by the South African National Biodiversity Institute and we thank Andrew Hankey of the Walter Sisulu National Botanical Gardens for his assistance in this regard. Diana Wall of Museum Africa provided access to the Fourie Collection and other poisoned arrows. The financial assistance of the National Research Foundation of South Africa towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the National Research Foundation.

Authors' contributions

J.B. and M.L. conceptualised the project; M.W. and D.K. prepared and ran the samples; V.M. interpreted the results; M.W., J.B. and V.M. wrote the paper; D.K. and M.W. prepared the figures; D.K., L.W. and M.L. provided academic input; and L.W. and L.P. provided conceptual input and read the draft manuscript.

References

1. Schapera I. Bushman arrow poisons. Bantu Studies. 1925;2:199–214. https://doi.org/10.1080/02561751.1923.9676184
2. Lee R. The !Kung San: Men, women and work in a foraging society. Cambridge: Cambridge University Press; 1979.
3. Marshall-Thomas E. The old way: A story of the first people. New York: Farrar, Straus and Giroux; 2006.
4. Bradfield J, Wadley L, Lombard M. Southern African arrow poison recipes, their ingredients and implications for Stone Age archaeology. South Afr Humanities. 2015;27:29–66.
5. Wadley L, Trower G, Backwell L, d'Errico F. Traditional glue, adhesive and poison used for composite weapons by Ju/'hoan San in Nyae Nyae, Namibia. Implications for the evolution of hunting equipment in prehistory. PLoS ONE. 2015;10(10), e0140269. https://doi.org/10.1371/journal.pone.0140269
6. Chaboo C, Biesele M, Hitchcock R, Weeks A. Beetle and plant arrow poisons of the Ju|'hoan and Hai||om San peoples of Namibia (Insecta, Coleoptera, Chrysomelidae, Plantae, Anacardiaceae, Apocynaceae, Burseraceae). ZooKeys. 2016;558:9–54. https://doi.org/10.3897/zookeys.558.5957
7. Mebs D, Brüning F, Pfaff N, Neuwinger H. Preliminary studies on the chemical properties of the toxic principle from *Diamphidia nigro-ornata* larvae, a source of Bushmen arrow poison. J Ethnopharmacol. 1982;6:1–11. https://doi.org/10.1016/0378-8741(82)90068-X

8. Woollard J, Fuhrman F, Mosher H. The Bushman arrow toxin, *Diamphidia* toxin: Isolated from pupae of *Diamphidia nigro-ornata*. Toxicon. 1984;22:937–946. https://doi.org/10.1016/0041-0101(84)90185-5

9. Kao C, Salwen M, Hu S, Pitter H, Woollard J. *Diamphidia* toxin, the Bushman's arrow poison: Possible mechanism of prey killing. Toxicon. 1989;27:1351–1366. https://doi.org/10.1016/0041-0101(89)90067-6

10. Jacobsen T, Sand O, Bjøro T, Karlsenz H, Iversen J-G. Effect of *Diamphidia* toxin, a Bushman arrow poison, on ionic permeability in nucleated cells. Toxicon. 1990;28:435–444. https://doi.org/10.1016/0041-0101(90)90082-I

11. Watt J, Breyer-Brandwijk M. The medicinal and poisonous plants of southern and eastern Africa. London: Livingstone Press; 1962.

12. Neuwinger H. African ethnobotany: Poisons and drugs: Chemistry, pharmacology, toxicology. New York: Chapman & Hall; 1996.

13. Marshall L. !Kung Bushman bands. Africa: Journal of the International Institute. 1960;30:325–355. https://doi.org/10.2307/1157596

14. Wadley L. The invisible meat providers: Women in the Stone Age of South Africa. In: Kent S, editor. Gender in African prehistory. Altamira: Walnus Creek; 1998. p. 69–81.

15. d'Errico F, Backwell L, Villa P, Degano I, Lucejko J, Bamford M, et al. Early evidence of San material culture represented by organic artefacts at Border Cave, South Africa. Proc Natl Acad Sci USA. 2012;109:13214–13219. https://doi.org/10.1073/pnas.1204213109

16. Lombard M, Haidle M. Thinking a bow-and-arrow set: Cognitive implications of Middle Stone Age bow and stone-tipped arrow technology. Camb Archaeol J. 2012;22:237–264. https://doi.org/10.1017/S095977431200025X

17. Williams V, Burke A, Lombard M. Throwing spears and shooting arrows: Preliminary results of a pilot neuroarchaeological study. S Afr Archaeo Bull. 2014;69:199–207.

18. Robbins L, Campbell A, Brook G, Murphy M, Hitchcock R. The antiquity of the bow and arrow in the Kalahari Desert: Bone points from White Paintings Rock Shelter, Botswana. J Afr Archaeol. 2012;10:7–20. https://doi.org/10.3213/2191-5784-10211

19. Backwell L, d'Errico F, Wadley L. Middle Stone Age bone tools from the Howiesons Poort layers, Sibudu Cave, South Africa. J Archaeol Sci. 2008;35:1566–1580. https://doi.org/10.1016/j.jas.2007.11.006

20. Lombard M, Phillipson L. Indications of bow and stone-tipped arrow use 64 000 years ago in KwaZulu-Natal, South Africa. Antiquity. 2010;84:635–648. https://doi.org/10.1017/S0003598X00100134

21. Lombard M. Quartz-tipped arrows older than 60 ka: Further use-trace evidence from Sibudu, KwaZulu-Natal, South Africa. J. Archaeol Sci. 2011; 38: 1918–1930. https://doi.org/10.1016/j.jas.2011.04.001

22. Wadley L, Mohapi M. A segment is not a monolith: Evidence from the Howiesons Poort of Sibudu, South Africa. J Archaeol Sci. 2008;35:2594–2605. https://doi.org/10.1016/j.jas.2008.04.017

23. Pargeter J, Shea J, Utting B. Quartz backed tools as arrowheads and hand-cast spearheads: Hunting experiments and macro-fracture analysis. J Archaeol Sci. 2016;73:145–157. http://dx.doi.org/10.1016/j.jas.2016.08.001

24. Wadley L, Sievers C, Bamford M, Goldberg P, Berna F, Miller C. Middle Stone Age bedding construction and settlement patterns at Sibudu, South Africa. Science. 2011;334:1388 – 1391. https://doi.org/10.1126/science.1213317

25. Borgia V, Carlin M, Crezzini J. Poison, plants and palaeolithic hunters. An analytical method to investigate the presence of plant poison on archaeological artefact. Quatern Int. 2017;427(B):94–103. https://doi.org/10.1016/j.quaint.2015.12.025

26. Van Wyk B-E, Van Heerden F, Van Oudshoorn B. Poisonous plants of South Africa. Pretoria: Briza Publications; 2002.

27. Taiz L, Zeiger E. Plant physiology. Sunderland, MA: Sinauer Associates; 2010.

28. Bisset N. Arrow and dart poisons. J Ethnopharmacol. 1989;25:1–41. https://doi.org/10.1016/0378-8741(89)90043-3

29. Van Wyk B-E, Van Oudtshoorn B, Gericke N. Medicinal plants of South Africa. Pretoria: Briza Publications; 1997.

30. Philippe G, Angenot L. Recent developments in the field of arrow and dart poisons. J Ethnopharmacol. 2005;100:85–91. https://doi.org/10.1016/j.jep.2005.05.022

31. Shaw M, Woolley P, Rae F. Bushmen arrow poisons. Cimbebasia. 1963;7:2–41.

32. Langejans GHJ. Remains of the day – Preservation of organic micro-residues on stone tools. J Archaeol Sci. 2010;37:971–985. https://doi.org/10.1016/j.jas.2009.11.030

33. Fourie L. Preliminary notes on certain customs of the Hei-//om Bushmen. Journal of the South West Africa Scientific Society. 1926;1:49–63.

34. Fourie L. The Bushmen of South West Africa. In: Hahn C, editor. The native tribes of South West Africa. London: Frank Cass; 1928. p. 81–105.

35. Wanless A. The silence of colonial melancholy: The Fourie collection of Khoisan ethnologica [PhD thesis]. Johannesburg: University of the Witwatersrand; 2007.

36. Bradfield J. Macrofractures on bone-tipped arrows: Analysis of hunter-gatherer arrows in the Fourie collection from Namibia. Antiquity. 2012;86:1179–1191. https://doi.org/10.1017/S0003598X0004833X

37. Pavarini D, Pavarini S, Niehues M, Lopes N. Exogenous influences on plant secondary metabolite levels. Anim Feed Sci Tech. 2012;176:5–16. https://doi.org/10.1016/j.anifeedsci.2012.07.002