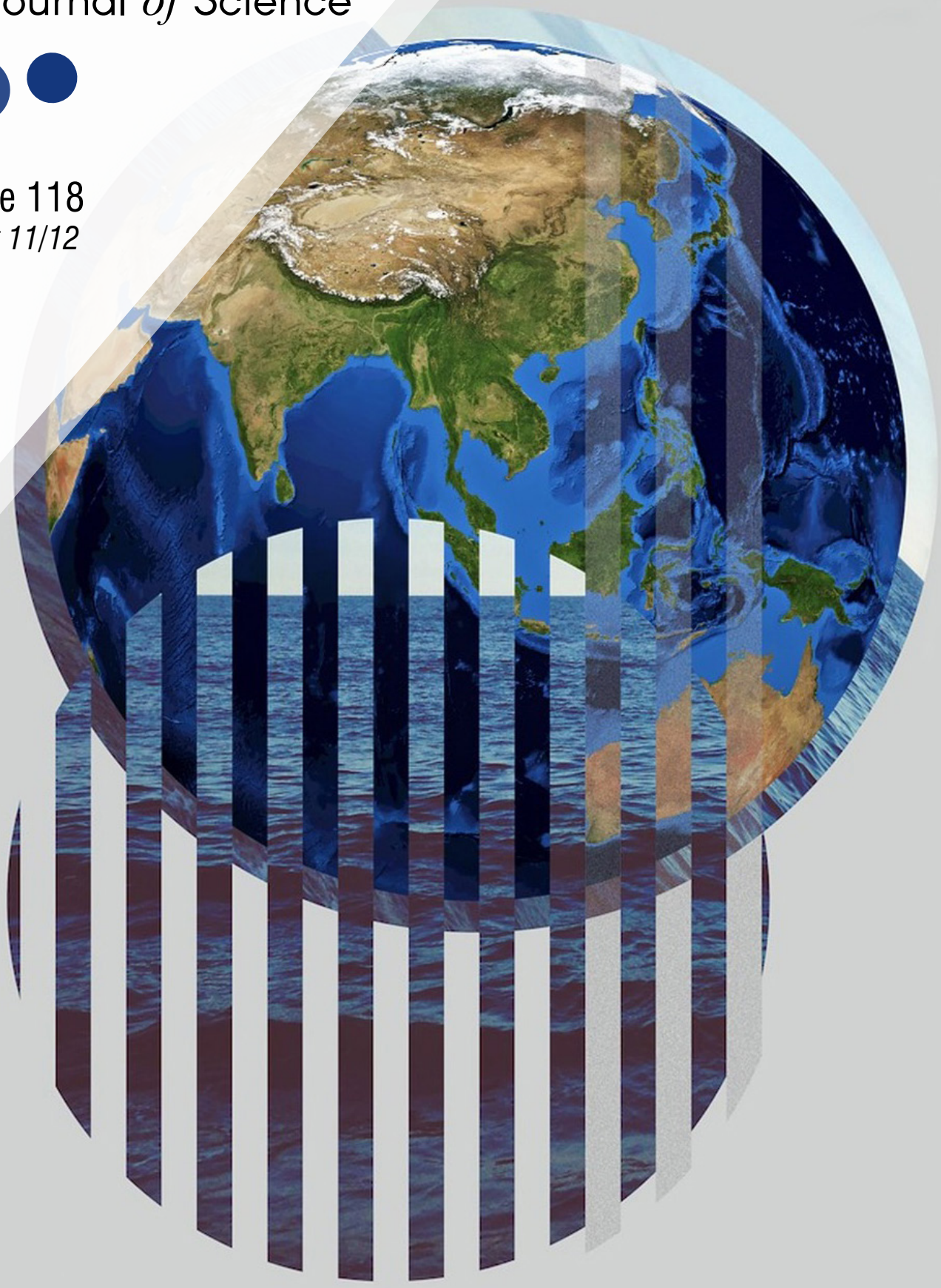
South African
Journal *of* Science

volume 118
number 11/12



Review of South African doctoral qualifications

Metabarcoding of marine zooplankton

Planetary health, human health and climate change

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Volume 118
Number 11/12
November/December 2022

South African Journal of Science

eISSN: 1996-7489

Leader

Book Reviews

Commentaries

Review Articles

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

Design and layout
Elzahn Swarts
E: swarts.elzahn@gmail.com

Correspondence and enquiries
sajs@assaf.org.za

Submissions
Submissions should be made at www.sajs.co.za

Research Articles

Research Letters

Cover caption
Ocean sustainable development, sustainability, and inclusivity are features of the Operation Phakisa 'Unlocking the Ocean Economy' Programme which strongly align it to a blue economy model (p. 11). How marine geophysics provides new insights into a Marine Protected Area (p. 72) and metabarcoding of zooplankton to indicate marine ecosystem status (p. 81) are the subjects of other articles on ocean health in this issue which focuses on climate change, and planetary and human health.

Inclusion through writing

On 5 December 2022, the Academy of Science of South Africa (ASSAf) will be hosting a session entitled 'Promoting social justice through accessibility of language in science' at the World Science Forum. Our journal is proud to be chairing this event, which coincides with the launch of our revamped language and style guide – what we term our Inclusive Language Policy.

This policy, which represents a more explicit codification of previous policies for our authors, is not a substantial shift in direction for the journal, but underlies and emphasises our commitment to the journal as a site for academic excellence and as much participation as possible by a range of scientists and scholars, mostly from our continent. There are two main features to the policy. First, we reiterate the importance of accessible, clear language across the board. For example, in keeping with many other scholarly journals worldwide, we encourage authors to avoid unnecessary jargon, and write in the active voice, in order to make their work easier to read. Second, we address particular issues in writing in an inclusive way, focusing on preferred terminology. Here, for example, we give some guidelines on reporting on race and ethnicity, sex and gender, and on disability, illness and impairment. We recognise in our policy that some terms and usages may be contested, and we leave space for two important contributors to what we hope will be a living and, where appropriate, changing document. First, we allow authors who have good reason to use language in different ways from those we suggest to contact us and provide reasons for this usage for us to consider. An example of this could be the preference in writing about neurodiversity for what is termed 'identity first' language rather than the 'people first' language we suggest in general for writing about disability, as the 'identity first' language is central to some approaches in the neurodiversity field. Second, we welcome discussion at a general level about the policy and its components.

Our wish to make language practices at our journal as inclusive as possible is not unique internationally, and is in line with many debates and positions put forward at the recent World Conference on Research Integrity held in Cape Town in 2022. It is also in keeping with the evolving Cape Town Statement which was discussed at the meeting, and to the Global Code of Conduct for research partnerships. Questions of fairness of access are key to all these endeavours. Language can be used to exclude and to divide; it is our responsibility as researchers to use language as far as we are able to include and to communicate respect.

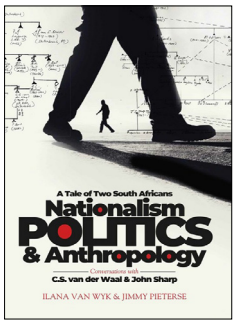
As uncontroversial as these views may appear on the surface, there are in fact a number of anxieties about accessible language which we would like to address. Possibly the most fundamental of these is the worry that using simple language will cause us to lower the standard of scientific communication. This is a worry which some express in a context where many dangerous and baseless claims are made and propagated through the Internet, and where the integrity and competence of good science and scientists are under attack. As a journal, we believe as strongly as ever in the importance of rigorous peer review, and we agree that there is a crucial difference between well-informed, evidence-based opinion and opinion based on untested or unfounded ideas. It is crucial that the standards of science are maintained, and all researchers have a role to play in emphasising the differences between rigorous work and populist declarations. Requiring scientific rigour, though, is not the same as requiring unnecessarily confusing writing and jargon. At times, scientists must use jargon in order to communicate with specialist peers, and at times we also use terms which may differ in specificity from how they are used in everyday talk – for example, there may be world of difference between somebody describing themselves as 'depressed' when they are writing a memoir, and the number of people a psychiatric epidemiologist would want to classify as 'depressed' when trying to assess the gap between the number of those requiring psychiatric treatments for depression and those actually receiving such treatments. As the science communication specialist Marina Joubert puts it, there is a big difference between 'dumbing down' (which is the last thing we want in good scientific communication) and 'clearing up' (which is something for which we strive) (https://www.scidev.net/asia-pacific/practical-guides/how-to-write-about-your-science-1/). Globally, for most people writing up their research in English, and most people reading such research, English is not their first language. The reasons for this go way beyond the global communication of science, and speak to the complex politics of the dominance of English internationally, which links to colonial and exclusionary histories. It is our responsibility to be as inclusive as we can to all who can contribute to research. One issue for us to explore, for example, is publishing abstracts in different languages. The issue of exclusionary language, though, extends beyond questions of whether people are writing and reading within their native languages. Even within the English-speaking world, there have been traditions of writing which make text unnecessarily obscure and imprecise. For example, some of us have been taught to use agentless passive constructions in our work (for example, 'It is well established'; 'It is believed'). These constructions obscure the basis on which claims are made. As researchers, we need to know who believes what, for example.

A second concern about our language policy may revolve around worries about the emotional and political sensitivities around terms used for race, gender, disability, and so on. It is indeed true that many people feel uncomfortable about issues of difference, and hence some may wish to resist what they view as attempts to police usage about difference. Some of the people we have consulted about our policy have expressed concerns about what they have termed 'political correctness'. The term 'political correctness' is in itself imprecise, but often links to anxieties that people and organisations (and in this case, our journal) may be pandering to a particular political constituency in a way that is performative rather than helpful for social cohesion or scientific rigour. Although we recognise that language is always linked to politics (this cannot be avoided), we believe that our policy helps clarify usage and works towards our (admittedly political but also broader) goals of inclusion, rather than to exclude and divide.

We envision that our policy will be helpful to authors and readers. As with all other aspects of the work of our journal, we welcome feedback and debate. It is in the nature of science that all in the science community must be open to new evidence and better arguments. We believe our policy is a clear marker of where we stand as a journal.

HOW TO CITE:
Swartz L, Fick L. Inclusion through writing. S Afr J Sci. 2022;118(11/12), Art. #15163. https://doi.org/10.17159/sajs.2022/15163

BOOK TITLE:
Nationalism, politics and anthropology: A tale of two South Africans: Conversations with C.S. van der Waal and John Sharp



AUTHORS :
Ilana van Wyk and Jimmy Pieterse

ISBN:
9789956552771 (paperback, 282 pp)

PUBLISHER:
Langaa RPCIG, Bamenda, Cameroon; GBP27

PUBLISHED:
2022

REVIEWER:
Rebecca Hodes[1]

AFFILIATION:
[1]Centre for Sexualities, AIDS & Gender, University of Pretoria, Pretoria, South Africa

EMAIL:
rebecca.hodes@up.ac.za

HOW TO CITE:
Hodes R. *A Tale of Two South Africans* – A review. S Afr J Sci. 2022;118(11/12), Art. #15090. https://doi.org/10.17159/sajs.2022/15090

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
30 November 2022

A Tale of Two South Africans – A review

The *Tale of Two South Africans: Nationalism, Politics & Anthropology: Conversations with C.S. van der Waal & John Sharp* is at once a biography and a meticulous intellectual ethnography. Written by Ilana Van Wyk, Associate Professor in Anthropology at the Department of Sociology and Social Anthropology at Stellenbosch University, and by Jimmy Pieterse, Doctor of Anthropology at the University of Pretoria, this book focuses on the history of anthropology in South Africa, through the experiences of two key anthropologists in the late 20th and early 21st centuries: C.S. 'Kees' van der Waal and John Sharp, both of whose careers spanned decades of apartheid rule as well as the first decades of democracy in South Africa.

The book emanated from a project to explore the knowledge of older anthropologists in Africa, who reach the compulsory retirement age of 65, often at the height of their academic careers and their power as scholars. As Francis Nyamnjoh, the principal convenor of this project, writes: 'The socio-political context of knowledge production, circulation and transmission should be systematically factored into anthropological analyses, theory building, and methodologies' (p. xiv). This is precisely what the book does in exploring anthropology's history in South Africa, particularly its contribution to debates about disciplinary approaches and theoretical bases.

The book examines the discipline of anthropology in the apartheid years through a focus on the establishment of *volkekunde*, which assigned 'an overwhelming explanatory power to the phenomenon of ethnicity, which it conceiv[ed] in the narrowest, most rigid terms possible' – terms that echoed apartheid racial and ethnic definitions.[1] The book traces the rise of *volkekunde*, which can be found in a German tradition that flourished before and during the Second World War and which centred *ethnos* as a core principle. Rather than simply study the discipline from the outside, one of the two key protagonists in the book, C.S. 'Kees' van der Waal, was trained in *volkekunde* at the University of Pretoria by the 'discipline's' most imminent disciples, P. J. Coertze and his son R. D. Coertze.

In conversation with van Wyk, the book traces van der Waal's eventual alienation from *volkekunde*:

> *I had to concede that the differences between Black and white cultural and technical performance were due to the differential access that each population had to education and capital. There was thus no justification for separate development or apartheid. My assumption that there were essential differences, a racist assumption that undergirded volkekunde, changed in that one moment. It was a massive and lasting paradigm shift.* (p. 126)

The book's second protagonist, with Pieterse as interlocutor, is the social anthropologist John Sharp. In contrast with van Wyk's conversations with van der Waal about his academic development, his early and later academic career, and his eventual and powerful break from *volkekunde*, Sharp's intellectual journey was starkly different. He was centrally involved in the intellectual project to critique apartheid through anthropological research. After the Universities of Cape Town and Cambridge, Sharp returned to South Africa to critique and dismantle the system from within. As an academic at the University of Cape Town, Sharp was central to intellectual shifts in the anthropology department, from 'Radcliffe-Brownian structural functionalism and Lévi-Straussian structuralism as the predominant paradigm in the anthropology of the 1980s' to *exposé* anthropology, which sought to expose the ways in which concepts such as culture were abused to legitimate the apartheid regime (p. 106). In addition to writing about these concepts, social anthropologists 'also focused on exposing the injustices of apartheid through detailed ethnographies of change and social production (Gordon and Spiegel 1993)'.[2]

Rather than position *volkekunde* as solely an alien project, wholly distinct from social anthropology and in the thrall of apartheid and Afrikaner nationalism, van Wyk and Pieterse seek to track *volkekunde's* emergence and evolution within the South African academy. While avowing the political complicity of *volkekunde* with apartheid, as well as the poverty of the 'discipline's' theoretical framework, van Wyk and Pieterse, as well as Sharp and van der Waal as their interlocuters, do the more difficult work of exploring the (wrong-headed) rationale for *volkekunde*, and seek to trace its emergence and evolution in South Africa, including the alignment of a leading *volkekundige* with the *Kolege ya Bana ba Afrika*, and his self-styled attempts to make the *Kolege* a constituent college of the University of Pretoria as a way to build its educational offerings to the same standards as those offered to white, Afrikaans students.

The text also explores Sharp's perhaps politically unpopular doubts about presenting *volkekunde* and social anthropology as 'two separate developments' in South Africa's intellectual history (p. 170). Particularly in the first half of the 20th century, the book provides examples of leading *volkekundiges* 'addressing the same sorts of questions as their social anthropological contemporaries and arriving at answers that were in certain respects not poles apart'. Sharp states that attempts to represent *volkekundiges* as 'another species of being' is 'to simplify a complex reality' (p. 171), and the book explores uncomfortable parallels between the two 'disciplines'.

Aside from these fascinating discussions and intellectual pursuits, what the text offers is a history of anthropology in South Africa from those who have lived it. It will be of interest to anyone working in intellectual history, as well as in ethnographic biography and epistemology.

References

1. Magubane B. The "Xhosa" in town, revisited urban social anthropology: A failure of method and theory. Am Anthropol. 1973;75(5):1701–1715. p. 1714. https://doi.org/10.1525/aa.1973.75.5.02a00310
2. Gordon RJ, Spiegel AD. Southern Africa revisited. Annu Rev Anthropol. 1993;22(1):83–105. https://doi.org/10.1146/annurev.an.22.100193.000503

BOOK TITLE:
Transformer: The deep chemistry of life and death

AUTHOR:
Nick Lane

ISBN:
9781788160544 (hardback, 400 pp; GBP25)
9781782834502 (ebook; GBP21)
9781782839989 (audiobook; GBP20)

PUBLISHER:
Profile Books, London, UK

PUBLISHED:
2022

REVIEWER:
Wieland Gevers[1]

AFFILIATION:
[1]Emeritus Professor of Medical Biochemistry, University of Cape Town, Cape Town, South Africa

EMAIL:
wieland.gevers@uct.ac.za

HOW TO CITE:
Gevers W. Not afraid to tackle most of the big problems in biology in one book. S Afr J Sci. 2021;118(11/12), Art. #15045. https://doi.org/10.17159/sajs.2022/15045

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
30 November 2022

Not afraid to tackle most of the big problems in biology in one book

It is a truism that the British genius is their wondrous way with words. From Chaucer and Shakespeare, they have led most of the other nations, even the French, in developing and spreading their language which has now become the world's primary intellectual commonspeak. Now we have an English author who is not (yet) a big name in biological science but has written yet another bold and clear book about nothing less than, inter alia, the origin of life, biological energy generation and utilisation, the emergence of species, ageing, cancer and consciousness. And his well-chosen words are worth noting carefully, because the author is very active and well published in many of the fields he covers, is an avid and perceptive reader of the literature, and is clearly a superb teacher, research innovator and communicator.

The author is also bold because he dares to think that the average intelligent citizen can understand the necessary physics, chemistry and biology provided enough trouble is taken to use metaphors and diagrams to make reactions and pathway schemes comprehensible, plus a good deal of repetition, humour and colloquialism to sweeten the pill. The writing in the book is so articulate and the unfolding narrative so ambitious that one is carried along helter-skelter, and the intermittently rough journey is made tolerable by entertaining and alternatively moving and alarming stories about individual scientific contributions by powerful personalities – in each case the actual experiments that revealed the breakthroughs are beautifully described and the generosity or skullduggery revealed. All in all, a first-rate intellectual adventure – even if I fear that not all of the bright ideas may turn out to be strictly true in the long run.

Just as a few top-down causation rebels wish to overturn the 'Modern Synthesis' of random mutation-driven natural selection, Lane would like everybody to turn away from the current genome-centric view of functioning living beings. Instead, he maintains that it is the flow of energy and not the state of the genome that determines whether a cell or an organism lives or dies. The machinery and the physicochemistry of this energy flow operate on time scales that extend from a moment to a lifetime, and it has to work over a huge range of rate requirements – it is composed of enzymes and membranes, metabolites in the form of substrates and products, and it *includes* the contributions of the two kinds of genomes in cells (nuclear and mitochondrial) but that involvement is not the whole story, far from it. Lane rejects the idea of accumulated mutations as the generalised cause of ageing, but rather thinks it is failing or aberrant mitochondrial respiration that resets metabolism on the path to senescence. He thinks the same is true for the origins of cancers. The detail-devil lies in the reasons why respiration is affected, and Lane has some ideas on this but cannot give full chapter and verse.

Two of the three biggest heroes of Lane's story are Sir Hans Krebs (this reviewer's own PhD supervisor) and Dr Otto Warburg, the first for his discovery of his eponymous biochemical cycle, which he thought was the irreversible final common pathway of the respiratory oxidation of most foodstuffs, and the second for his observation of the 'Warburg effect', namely that cancer cells regularly display impaired respiration and instead rapidly break down glucose via glycolysis even when oxygen is freely available, something which normal cells do not do (in turn a prior finding of Louis Pasteur known as the 'Pasteur effect'). Lane uses a wealth of more recent work which has shown that the Krebs Cycle is in fact fully reversible, with alternative enzymes or bypass shuttles being available to overcome the steps in the cycle that Krebs thought were irreversible. In the 'Warburg effect', the cycle in reverse becomes reductive instead of oxidative, biosynthetic instead of catabolic, and a potent engine of the growth of cells, provided that an adequate alternative energy supply is available. The contrasting roles of the two nicotinamide coenzymes, NAD and NADP, are respectively the main cellular oxidising and reducing agents of the forward and reverse cycles. Some Krebs Cycle intermediates are key signalling molecules to set in train multiple pathways affecting growth and inflammation in both cancer and ageing. (In the last-mentioned context, it is curious that Lane does not once mention the well-known operation of the same 'Warburg effect' in activated immune cells such as M1 or pro-inflammatory macrophages, a topic of great current interest in immunology.) The key place of the amino acid glutamine in supporting the growth of cancer cells and in linking their expansion to muscle wasting is, however, emphasised.

The book does not fully recognise, as a potential third great hero of all this biological machinery, Fritz Lipmann, who developed the concept of cellular work of all kinds being done by the splitting of so-called 'high-energy' bonds, meaning that chemical 'free energy' is released that is coupled mechanistically to anabolic reactions (e.g. the biosynthesis of macromolecules like proteins and nucleic acids) or processes (e.g. muscle contraction or ion pumping against gradients in nerve cells). The oxidation of foodstuffs in cells is primarily aimed at forming the predominant 'high-energy' molecule, adenosine triphosphate or ATP. The reason that Lane seems to downgrade Lipmann's emphasis on ATP as the general 'energy currency' of all living cells and organisms, is his (in my view not fully justified) pre-occupation with electricity as the hallmark of life. This leads to *his* choice of third hero as being the maverick scientist Peter Mitchell, whose Nobel Prize-winning work revealed that respiration in mitochondria, driven mostly by the oxidations of the forward Krebs Cycle, actually leads to the formation of ATP through the creation of a powerful 'proton-motive force' (steep proton gradient) as well as electrical potential difference across the mitochondrial membranes. It is the combination of these two factors which causes the positively charged hydrogen ions (protons) to re-enter the inner mitochondrial space, which they can only do via a special membrane-embedded enzyme that makes ATP in profusion, in obligatory coupling to the proton flow. Lane (somewhat disconcertingly, even alarmingly) calculates that the 150–200 millivolts across the 6-nm-thick membrane, as experienced by a molecule present in it, is equivalent to a bolt of lightning!

Lane pursues these ideas (some of which he has also covered extensively in several previous books) to postulate the likely origin of all life in inorganic chemical reactions near deep-water volcanic vents, involving hydrogen and carbon dioxide and sulphur compounds. Protocells formed with partitions that began to be able to develop electrical charge; chemical syntheses of early life-molecules took place without enzymes; primitive RNAs began to code for some of the biosyntheses and to act like enzymes, and gradually the earliest unicellular forms of life appeared. The Krebs Cycle, when it finally worked as a system catalysed by separate enzymes and coenzymes, was all in reverse, a biosynthetic highway. As photosynthesis (also described entertainingly and informatively in a separate chapter by Lane) generated oxygen and this gas accumulated, organisms began to use the forward Krebs Cycle to form ATP but this was largely restricted to the multicellular animals which appeared on earth during and after the 'Cambrian explosion', where some organs needed to do work at high rates and others could concentrate on biosynthesis – so the Cycle became facultative, sometimes going one way and sometimes another, even splitting up sometimes to go in two different directions. A versatile metaphoric traffic circle at the very heart of life. The late Sir Hans would indeed be amazed.

Lane's dominant electrical/biophysical emphasis persists throughout his book (even to a theory of consciousness which questions conventional interpretations of the electro-encephalogram). Thus, he uses the term 'flux capacitator' in one chapter dealing inter alia with fascinating insights into the roles of multiply mutated mitochondrial DNAs in sexual reproduction and possibly in speciation, building on the pioneering work of Douglas Wallace. He devotes much attention to homeostatic mechanisms within the rather dangerous-sounding mitochondrial respiratory environment, involving tight control of the formation of the essential but dangerous 'reactive oxygen species' that are formed mainly by the first complex of the respiratory chain, balanced by the formation of ATP when the chain operates through its full series of components to reduce oxygen to water.

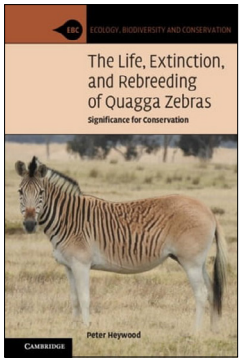
The overall title of the book, the word 'Transformer' (usually defined as a 'device that transfers electric energy from one alternating-current circuit to one or more other circuits, either increasing (stepping up) or reducing (stepping down) the voltage') is not very clearly justified by the author if he meant it to signify the electrical rather than the more general meaning of the term.

This reviewer is convinced of Lane's 'electrical' approach to some degree, but is puzzled by the omission of any mention of alternative or additional mechanisms of energy regulation in living cells. For example, Krebs himself proposed an enzymatic 'flux capacitator' in the form of the equilibrium reaction catalysed in cells by the very active adenylate kinase (myokinase), in which a small fall in ATP concentration results in a large increase in AMP, an allosteric effector that can affect the rates of many key enzymes in glycolysis and the Krebs Cycle. AMP is also a strong activator of a very important protein kinase (AMP-activated protein kinase) with extensive pleiotropic effects on energy metabolism. This kind of selectivity in adducing evidence makes one think that the chemistry/biochemistry of cells and organisms has in the distant past been, and still is, a lot more important in making everything biological happen than Lane gives it credit for.

Nevertheless, all in all, this book is very important because it presents a wealth of ideas coherently drawn together and extending to a wide range of biological phenomena. It is an example of how a full-length exposure in a book of a scientist's 'idea-world' can be more creative and influential than a whole series of journal papers. The book will also be very useful in undergraduate and postgraduate teaching, and in widening the horizons of biologists of all stripes and persuasions. Highly recommended.

The ecology, extinction and resurrection of the quagga

BOOK TITLE:
The life, extinction, and rebreeding of quagga zebras: Significance for conservation



AUTHOR:
Peter Heywood

ISBN:
9781108831604 (hardcover, GBP85, 230 pp)
9781108926911 (paperback, GBP40, 230 pp)

PUBLISHER:
Cambridge University Press, Cambridge, UK

PUBLISHED:
2022

REVIEWER:
Brian W. van Wilgen[1]

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

EMAIL:
bvanwilgen@sun.ac.za

HOW TO CITE:
van Wilgen BW. The ecology, extinction and resurrection of the quagga. S Afr J Sci. 2022;118(11/12), Art. #15176. https://doi.org/10.17159/sajs.2022/15176

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
30 November 2022

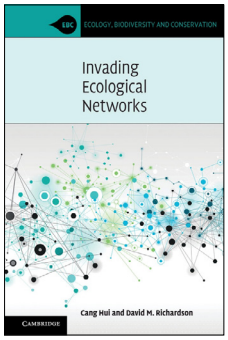
There are many global icons of extinction – the best known of which is the dodo (*Raphus cucullatus*), a flightless relative of pigeons that lived on Mauritius until the 17th century. In South Africa, our icon is undoubtedly the quagga (*Equus quagga quagga*), a subspecies of the plains zebra, which once occurred in astounding numbers across the karoo and grasslands of South Africa. It was hunted to extinction in the late 19th century, and the last living specimen died in a zoo in Amsterdam in 1883. This book is an account of where quaggas fit into the family of equines, their distinguishing features, their interactions with people, and the causes of their extinction. It deals further with the subsequent futile endeavours to locate unlikely survivors well into the 20th century, the later ongoing attempts to breed animals that resemble extinct quaggas, and the philosophical arguments for and against this rebreeding campaign.

The author, Peter Heywood, is a professor of biology at Brown University in the USA. His interest in quaggas began in 2006, and he has published several papers that cover biological and social perspectives (including art and poetry) relevant to quaggas. His interests have motivated him to examine almost all of the extant illustrations, paintings, photographs and museum specimens of quaggas, both in Europe and South Africa. The historical accounts of the quagga's discovery, description and, ultimately, their extinction make for interesting reading, but the material of most scientific interest is dealt with in the last two chapters on attempts to recreate the species, and on the place and value of such attempts in the modern world.

The book describes the key role played by the German-born taxidermist Reinhold Rau (1932–2006), who joined the staff of the South African (now Iziko) Museum in Cape Town in 1959. The museum held a quagga foal specimen that he re-mounted to meet his exacting standards, initiating his interests in quaggas, and later his passion to rebreed them. The major breakthrough came with the cloning and sequencing in 1984 of quagga DNA that Rau had managed to glean from museum specimens. The DNA revealed that the quagga was a subspecies of the extant plains zebra, and not a separate species as was believed at the time. This suggested that selective breeding to diminish striping on the legs and hindquarters could be used to produce animals that resembled quaggas. The programme began with selected zebras from Namibia and northern KwaZulu-Natal in 1986, and is ongoing. Today a sizeable population of 'Rau quaggas' (so named to distinguish them from true quaggas) occurs on several protected areas and farms within the quagga's historical range.

The final chapter focuses on the identity of rebred quaggas and the value of rebreeding for conservation. Biologists clearly recognise that the phenotypes (specimens that externally resemble quaggas) will never be true genotypes, with all of the behavioural and physiological features that distinguished the extinct subspecies, but they may have cultural or symbolic value. The question is asked as to whether the funds expended on trying to revive an extinct species should not rather be spent on saving extant species that face extinction right now. For example, the Selous' zebra (*Equus quagga selousi*) is another distinctive subspecies of the plains zebra with a small remnant population in Mozambique – if we act now, we could avoid having to attempt rebreeding it later – or losing it forever. These and other debates are very relevant to how we handle the accelerating sixth extinction. The book is an enjoyable account of this interesting case of demise and attempted resurrection, and will appeal to conservation scientists as well as wildlife enthusiasts.

BOOK TITLE:
Invading ecological networks



AUTHORS :
Cang Hui, David M. Richardson

ISBN:
9781108745963 (paperback, 400 pp)

PUBLISHER:
Cambridge University Press, Cambridge, UK; ZAR864

PUBLISHED:
2022

REVIEWER:
Kevin J. Duffy[1]

AFFILIATION:
[1]Institute of Systems Science, Durban University of Technology, Durban, South Africa

EMAIL:
kevind@dut.ac.za

HOW TO CITE:
Duffy K. Attacking the role of invasion in ecology: A holistic approach. S Afr J Sci. 2022;118(11/12), Art. #14693. https://doi.org/10.17159/sajs.2022/14693

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
30 November 2022

Attacking the role of invasion in ecology: A holistic approach

This book is a *tour de force* of invasion science and to some extent the entire science of understanding ecology through the interplay of ideas and empirical studies mediated by models. Ecological systems are clearly complex entities with interrelated components. The authors take the non-reductionist approach of studying ecology through the lenses of network analyses. In this they focus on biological invasions. While I was very aware of the relevance of invasion biology and especially that of alien invasions, I tended to view the latter as an important evil to be avoided. Hui and Richardson place invasion science firmly as a central player in current ecological systems as almost all are affected by the biological invasions brought on by the dominant presence of *Homo sapiens*.

The authors "view 'ecological networks' as entangled webs of distinct, interdependent, and complex biotic interactions among co-occurring species in a landscape". While this approach appears obvious, biological invasions are often researched in a straightforward progression of introduction, establishment, and spread. This book develops modern network and complexity science to consider invasions more holistically. For this approach, ecosystems are modelled as open adaptive networks with states governed by multiplayer eco-evolutionary games. The authors stress that it is not a textbook but rather a 'cloud atlas' to the study of invasion science, which I take to mean a guide to methods toward any successful understanding of the subject. This claim might appear somewhat ambitious or even arrogant, but they definitely make a strong claim to having provided an atlas for the study of invasion science.

Chapter 1 covers the key concepts basic to understanding biological invasions. Here the foundations are set as to why a more comprehensive approach is necessary. Chapter 2 introduces the importance of using network approaches for studying these systems. Described is an understanding of ecosystems as a group of entities in a type of game of survival. These entities exist in different states, but at the most basic, the individual strives to survive and pass on its genetic material. As there are many players, this requirement depends on each individual having strategies to optimise their survival when in competition or cooperation with others. The chapter maps out how these strategies can be studied as scenarios of multiplayer games and using network science. Using standard mathematical methods developed mostly around differential equations, but applied to the network, the stability of the system can then be assessed toward an understanding of the coexistence and invasions between species in an ecological system. Chapter 2 is titled 'Relentless Evolution' and the breadth of the methods is a relentless coverage.

Chapter 3 covers in what way networks assemble. A lot of knowledge has been developed in the past decades around how real-world networks evolve and function. For ecological systems, the chapter highlights the importance of three network types: competitive, antagonistic, and mutualistic. Mutualism tends toward nested networks and antagonism toward compartmentalised networks and this framework of network structure leads onto Chapter 4 where the dynamics that play out are described. Chapter 3 also covers in what ways networks emerge.

Chapter 4 describes how the stability of complexity networks can be applied to ecosystems and especially how invasions can disrupt stability driven dynamics. First, the authors point out the importance of viewing ecosystems as open adaptive systems rather than superorganisms to avoid the danger of applying evolutionary concepts to these superorganisms. This approach then allows for the application of a theoretical framework based on complex adaptive systems to consider ecosystems and their stability. Structural stability is considered by measuring the possible persistence of the system when removing or adding species. Invasions can disrupt ecosystem stability toward a new state or an unstable state and ultimately collapse. Large changes can result from small invasions if at some critical point. Chapter 4 offers much greater detail to the theoretical methods used to study these complex systems and a strong case is laid out to support the view of ecosystem networks as a 'jammed complex system self-organised at marginal instability due to the flow from biological invasions and resulting species extinctions'.

Chapter 5 goes into the difficulty of predicting future ecosystem states and dynamics with different possible ways to attempt prediction presented. The authors present a novel method to measure the short-term direction of an ecosystem, again using the dynamical systems analyses of differential equations. A counterintuitive result from this chapter is how rare species might control resilience of the system, with some evidence from actual ecosystems. I think this is overstated but am happy to be corrected by the authors. However, overall, this chapter describes how common and rare species can drive the system in various ways.

Chapter 6 discusses how biomes dominated by alien species emerge. This chapter also considers how spatial and temporal scales can affect network structure and stability and how propagation by dispersal affects the stability and dynamics of meta-networks. Chapter 7 summarises and pulls together a lot of the concepts of the book.

As I stated at the start, this book is a *tour de force* of the subject. However, as such, it is rather overwhelming. Many aspects of nonlinear mathematics and science are touched on. While I, over many years, have studied a lot of the mathematical theory presented, it is daunting to see it applied in so many ways to a particular, if complex, question in science. The authors themselves suggest at the end of Chapter 5 that the reader might be a bit giddy by this stage (as Alice in her Wonderland was). I certainly was. They suggest staying the course. For an ecological scientist this could be good advice. For the general reader, I would suggest picking out what they can from the many interesting ideas, and if giddy happily go back up 'the rabbit hole'. Without a doubt, the book clearly shows the importance of invasion biology for the planet and its inhabitants.

AUTHORS:
Fiona N. Mumoki[1]
Christian W.W. Pirk[1]
Abdullahi A. Yusuf[1]
Robin M. Crewe[1]

AFFILIATION:
[1]Social Insects Research Group, Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Fiona Mumoki

EMAIL:
fiona.mumoki@up.ac.za

HOW TO CITE:
Mumoki FN, Pirk CWW, Yusuf AA, Crewe RM. Anniversary of a beekeeper's discovery of thelytoky in Cape honey bees. S Afr J Sci. 2022;118(11/12), Art. #14126. https://doi.org/10.17159/sajs.2022/14126

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
G.W. Onions, thelytoky, Cape honey bees, laying workers, reproductive parasitism

FUNDING:
South African National Research Foundation (grant no. 129608)

PUBLISHED:
30 November 2022

Anniversary of a beekeeper's discovery of thelytoky in Cape honey bees

Significance:

- The laying workers of the Cape honey bee continue to negatively affect the South African beekeeping industry, with more losses suffered in the northern regions of the country.
- The reproductive parasites enter susceptible host colonies, activate their ovaries, and lay diploid eggs, leading to colony dwindling and collapse.
- Diploidy in eggs produced by unmated laying workers arises from thelytokous parthenogenesis, first discovered in honey bees by a hobbyist beekeeper.
- We examine the consequences of thelytokous parthenogenesis and outline what is being done to understand and limit the spread of the laying workers of the Cape honey bee.

Introduction

The existence of thelytoky in Cape honey bees was first documented in 1912 – making 2022 the 110th anniversary of this discovery. The significance of this discovery took a number of decades to be fully appreciated. The biology of sex determination in honey bees – where males are produced from unfertilised eggs while females arise from fertilised eggs – allows the reproductive female to control the sex ratio of her offspring. Hence, queens in honey bee colonies bias the sex ratio in favour of producing large numbers of females (that will be workers) and only producing males seasonally when needed. This gives rise to an exclusively female social group characterised by a single reproductive female (queen) and large numbers of non-reproductive females (workers). As workers are unable to mate, any offspring they produce should be male. Finding that Cape honey bee workers can produce female offspring was an anomaly that required extensive investigation to unravel its biological significance.

South Africa is home to two subspecies of honey bees. One is *Apis mellifera scutellata,* also known as the Savannah honey bee predominantly found in the northern parts of South Africa, extending northwards into various parts of East Africa. The second subspecies – *A. m. capensis* also known as the Cape honey bee – is native to the Western Cape region of South Africa in the Fynbos Biome. Separating these two subspecies is a natural and stable zone of introgression, restricting the naturogenic spread of the Cape honey bee outside of its native region (Figure 1A). *A. m. capensis* workers are usually darker than the typically yellow and black *A. m. scutellata*. Although colour alone is an unreliable feature to separate the subspecies, it has some utility. Unlike honey bee populations in the northern hemisphere, the majority of honey bees in South Africa are wild and not managed.

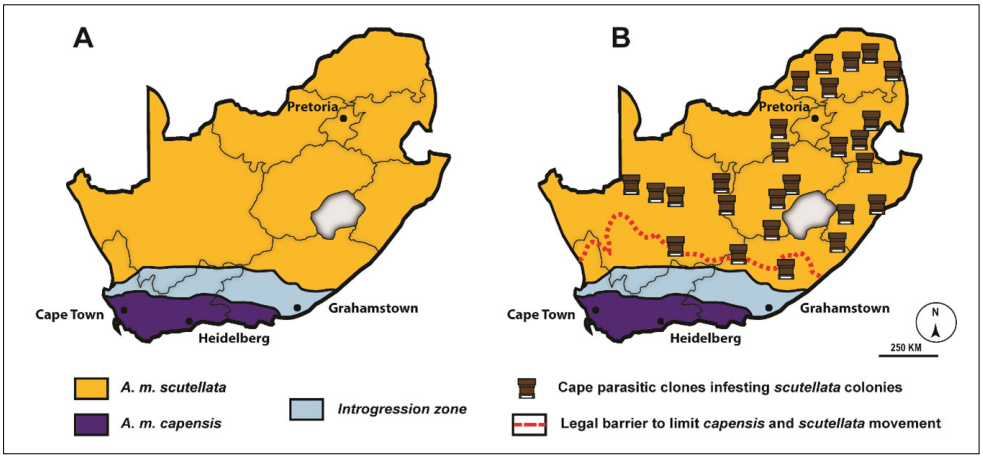


Figure 1: Map of South Africa showing the distribution of the two subspecies of honey bee – *Apis mellifera scutellata* (yellow) and *A. m. capensis* (purple) – (A) prior to 1990 and (B) after the anthropogenic establishment of parasitic laying workers (brown hives) after 1990. The red dotted line shows the legal barrier (Reg. 159/5 February 1993) enacted to restrict the movement of colonies of *A. m. capensis* and *A. m. scutellata* beyond the stable zone of introgression (light blue).

In 1993, beekeepers in South Africa were estimated to have lost 50 000–100 000 honey bee colonies, but not due to pests, diseases or pesticides, which are the usual causes of decline in honey bee populations in many parts of the world.[1] These beekeepers were facing a new challenge, from dark worker bees that were invading healthy but susceptible host colonies, laying unfertilised diploid eggs and producing other dark workers that were also reproductively active. The newly emerged workers would in turn seek out and invade other colonies, eventually leading to dwindling and collapse of the infested hosts, called the 'Capensis calamity'.[2] These laying workers have spread widely into the *A. m. scutellata* native range, with the spread attributed to their presence in managed honey bee colonies (brown hives, Figure 1B). Subsequently, Pirk et al.[3] showed that beekeepers in South Africa were

losing about 41% of their colonies annually, with a significant cause of this loss attributed to the presence of *A. m. capensis* parasitic workers.

This ability of Cape honey bee workers to lay eggs that develop into diploid (females) individuals had been described by a little-known hobbyist beekeeper George William Onions (Figure 2), in a landmark paper describing reproductive parasitism, published in the *Agricultural Journal of the Union of South Africa*.[4]

Figure 2: George William Onions (1866–1941), the hobbyist beekeeper who was among the first to make critical and detailed observations of reproductive parasitism by *Apis mellifera capensis* laying workers.

G. W. Onions

George William Onions was born in 1866 to William Onions and Miriam Lockhart in Cape Town, South Africa. He was the second of three children, though unfortunately when he was still young his mother and both siblings died of diphtheria. His father would soon pass away too, leaving the young Onions in the care of his aunt. At the age of 28, he married Jessie Elizabeth Massey and together they had seven children.

Not much is known about his formal education, though from accounts given by his daughters, Onions is said to have been mostly self-taught. By profession, Onions was an electrician, operating a business that installed electrical wiring in buildings. He was also a keen inventor and is said to have invented an X-ray machine and even made a small telescope, which he mounted on a stone and used to observe stars.

His biggest passion, however, was beekeeping and this was evident in his landmark paper 'South African "Fertile-Worker Bees"' published in 1912[4], in which he described himself as '*…a practical beekeeper and queen breeder who has always endeavored to verify in his own experience, the facts and theories of bee culture…*'. From his understanding of scientific bee literature, it seems that he was also a keen reader of this literature and curious about basic honey bee biology.

Indeed, Onions was a beekeeper who purely through observation and basic experimentation came to important discoveries, later verified by researchers using modern analytical techniques. Starting his observations in 1901[4] to 1913[5], he described how in 1901 he clipped the wings of a virgin queen, mistaking it for a mated queen due to the presence of hundreds of freshly laid eggs in the colony. He later found out that the eggs had been laid by workers of African origin that had infested his Italian honey bee colonies (Mr Onions daughter, Mrs E.J. Walton, confirmed that her father imported bees from the USA and Europe, in an effort to breed honey bee races more docile than *A. m. scutellata*). Onions described that the 'dark coloured' bees had the ability to lay eggs, even in the presence of a queen. In describing the life cycle of a colony infested by laying workers, he showed that these workers would seek out susceptible colonies that were small or queenless and activate their ovaries (verified through dissections) to lay eggs. The eggs laid would emerge as workers which – Onions observed – were not performing the typical colony maintenance tasks such as brood rearing, comb building and foraging and would instead become reproductively active. This would eventually lead to the death of the host queen, and dwindling and eventual collapse of the host colony.

Further, Onions recorded that the laying workers had a well-developed 'sperm sac', which was smaller but structurally similar to that of the queen. In 1913, only queens were known to have spermathecae, making this a very controversial claim. He went on to describe that the spermatheca of the Cape bees was empty (unlike that of a mated queen), and thus, these bees had produced workers from unfertilised eggs. This observation was in clear disagreement with the prevailing understanding of the time as described by the Polish theologian and bee scientist Johann Dzierżoń who in 1845 discovered that unfertilised eggs produced males and thus drones while fertilised eggs produced females and thus workers or queens (Figure 3A). By 1906, Dzierżoń's observations were accepted

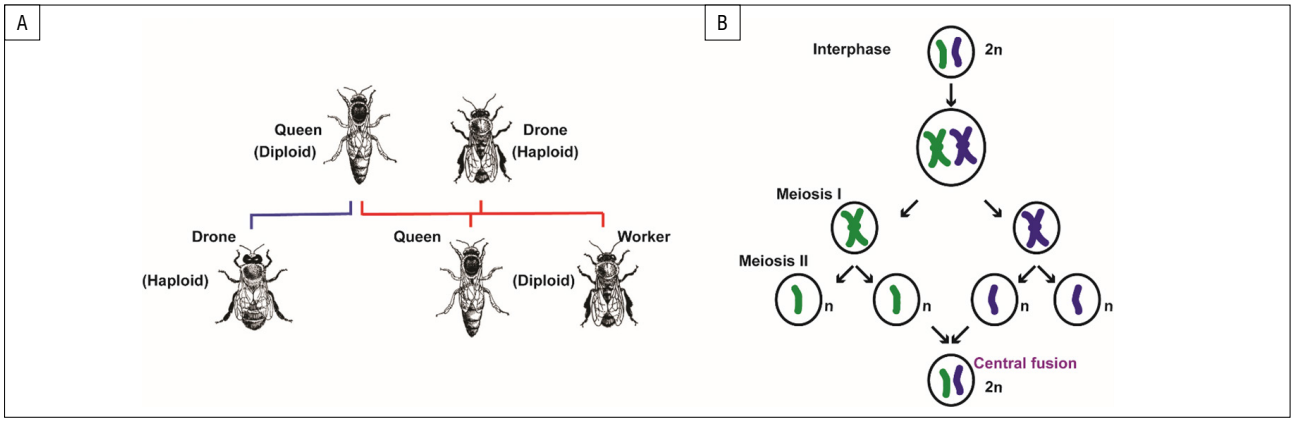


Figure 3: Sex determination by haplodiploidy as described by Johann Dzierżoń in 1845 where (A) unfertilised (haploid) eggs become drones while fertilised eggs become workers or queens. However, the diploid state can be restored in thelytokous parthenogenesis (B), through the fusion of the two central pronuclei in meiosis II, as happens in the laying workers of *Apis mellifera capensis*, leading to emergence of females from unfertilised eggs, thus being an exception to Dzierżoń's rule.

as the established mode of sex determination in honey bees, therefore requiring Onions to defend his observations that the reproductively active Cape honey bee workers were laying unfertilised eggs that led to the emergence of female bees. It would be more than 75 years before Savitri Verma and Friedrich Ruttner in 1983 explained how diploidy is restored in these bees, experimentally confirming Onion's observations.[6]

For the time being though, Onions had to contend with working with Rupert Jack, a Rhodesian government entomologist who took it upon himself to validate the veracity of Onions' incredible claims[7] given the fact that Onions was not an entomologist. One of the fiercest critics of the observations by Onions came from van Warmelo who refuted Onions' conclusions based on the findings of Dzierżoń and claimed that if the 'fertile workers' were producing workers and not drones, these fertile workers must have mated with drones.[8]

Prior to 1910

A previous attempt had been made to explain the curious behaviour of the laying workers of *A. m. capensis* by Sir Henry de Villiers who, as well as serving as the first Chief Justice of South Africa, was an avid beekeeper. In a study published in 1883, Lord de Villiers described his discovery of eggs laid in cells in a colony that had just recently moved into his hive, and that had caged its queen to prevent her from absconding. In further observations, he noticed that the adult bees emerging from these eggs were workers. Possibly based on Dzierżoń's work and given that there was no other queen in the colony, Lord de Villiers concluded that the eggs must have been laid by the caged host queen and passed through the holes of the cage to workers who then deposited them in the brood cells.[9]

1920–1960

A large portion of the bee research in South Africa at this time focused on attempts to breed a more docile race of the South African bees, through the importation of European bees. Much of this work was championed by A. E. Lundie (Bee Culture Laboratory of the Bureau of Entomology). Although not much research into the laying workers of *A. m. capensis* took place during this time, Lundie published an article confirming the ability of Cape honey bees to lay worker-destined eggs.[10]

It was R. H. Anderson (Agricultural Department, Plant Protection Research Institute of the Agricultural Research Council) who first reexplored the findings of Onions. Anderson reported that laying workers were present in *A. m. capensis* colonies, even in the presence of the queens.[11] Further, he showed that, following the removal of a queen, infighting would take place, after which some workers would quickly activate their ovaries and begin laying eggs. In validating Onions' findings, Anderson further showed that the laying *A. m. capensis* workers had activated ovaries with many ovarioles per ovary, a large spermatheca that contained no sperm and, very importantly, that the laying workers were producing female offspring parthenogenetically. This was 50 years after Onions' groundbreaking paper.

After 1970

Prior to 1970, beekeeping in South Africa was mainly done for the purpose of producing hive products such as honey, wax, and propolis. However, the industry underwent a major transition after 1970 with the realisation that there was a demand for pollination services that could be more lucrative than simply selling hive products. To achieve this, the beekeepers began moving their colonies from place to place – a process that greatly facilitated the spread of the *A. m. capensis* laying workers, leading to dwindling and collapse of the infested host colonies.

The Capensis calamity

In 1912, Onions reported his detailed observations describing the host-seeking behaviour of *A. m. capensis* laying workers, activation of ovaries and laying eggs and the dwindling and eventual collapse of infested colonies.[4] It would be more than 80 years before these laying workers would be recognised as a major problem for the beekeeping industry in South Africa, as described by Allsopp and Crewe[2]. The report followed a major outbreak of *A. m. capensis* laying workers that occurred in 1990–1991, in areas outside of the native range of *A. m. capensis* (Figure 1). Prior to this, there had been two other outbreaks in the 1980s as reported by Geoff Tribe and Martin Johannsmeier, although these outbreaks were on a small scale and easily contained. The 1991 outbreaks, however, involved a large number of *A. m. capensis* colonies moved outside their native region and also transport of *A. m. scutellata* colonies to the *A. m. capensis* native regions and back, greatly facilitating the spread of parasitic workers.[12]

Dietemann et al.[12] and Pirk et al.[3] showed that migratory beekeepers face higher colony losses due to the laying *A. m. capensis* parasites, and herein lies the crucial role that beekeepers can play in stopping the 'Capensis calamity'. As recommended by many researchers investigating the nature and impact of the 'Capensis calamity' on South African apiculture, beekeepers should only utilise locally caught bee colonies for beekeeping and should limit the migration of bees from place to place as migration increases the risk of exposure of colonies to infestation by parasitic workers. Further, the *A. m. capensis* honey bees should never be moved outside of their native range and into that of the *A. m. scutellata*.[13]

Huge economic losses sustained by beekeepers, the increased distribution of the parasitic laying workers and the potential threat to wild honey bee populations and to food security posed by the continuing spread of the *A. m. capensis* laying workers refocused the attention of researchers and other stakeholders of the beekeeping industry. More scientists shifted their attention to this issue, trying to understand the biology of the reproductive parasites, how these laying workers seek out susceptible colonies and evade detection by host workers, and why they are able to fully activate their ovaries and lay eggs, even in the presence of the host queen.

While most of the research prior to 1970 was based almost purely on field observations, the advent of more advanced scientific tools in gas chromatography (GC), microscopy and molecular genetics enabled researchers to address deeper questions needed to understand the biology of the *A. m. capensis* laying workers. In great collaborative and multidisciplinary efforts, scientists teamed up to work on various aspects of the biology of the laying workers, as reviewed by Mumoki et al.[14], confirming and extending the observations reported by Onions in 1912[4] and 1914[5].

Through backcross experiments, Verma and Ruttner[6] showed that the restoration of the diploid condition in *A. m. capensis* worker-laid eggs occurs through fusion of the two central pronuclei, in what is known as thelytokous parthenogenesis, thereby explaining the emergence of diploid workers from unfertilised eggs laid by unmated workers (Figure 3B). Further, using GC and GC-mass spectrometry, Crewe and Velthuis[15] and others[14] demonstrated that – in addition to the presence of a spermatheca, high ovariole number per ovary and an ability to rapidly activate their ovaries to lay eggs – the *A. m. capensis* laying workers were also producing queen-associated mandibular gland pheromones. These queen-associated signals are not just from the mandibular glands but also from the Dufour's and the tergal glands – a further indication of the involvement of multiglandular pheromone signals in establishing reproductive dominance. Various molecular studies[14] have shown that the laying Cape honey bee workers utilise queen-associated pathways in the biosynthesis of the fatty acid components of the mandibular gland pheromone. Ongoing work is focused on exploring the molecular mechanisms involved in the biosynthesis of the communication signals from the tergal and Dufour's glands.

Hepburn and Crewe[16] showed that while the principal mode of reproduction in *A. m. capensis* is thelytokous parthenogenesis, workers of this subspecies are also able to reproduce arrhenotokously (produce male offspring from unfertilised eggs) and that the contribution of worker reproduction within the South African honey bee population is significant. What governs the switch from arrhenotokous to thelytokous parthenogenesis has been the subject of ongoing research, with conflicting conclusions that will require use of functional characterisation of the genes involved using tools such as RNA*i* and CRISPR-Cas9.

Finally, as this groundbreaking discovery showed, the link between beekeepers and social insect researchers is essential and is a manifestation of 'citizen science' in action. Key aspects of the life of the reproductive parasite *A. m. capensis* were first and rather accurately described by a keen and curious beekeeper who took meticulous notes of his observations for 10 years before presenting his findings to the scientific community – findings that many scientists have built on over the years using various tools in genomics and chemical ecology, to understand the biology of the laying workers of *A. m. capensis.* Indeed, the Cape honey bee has provided us with a unique opportunity to witness the evolution of a social parasite in real time, providing us with significant insights into the evolution and regulation of the reproductive division of labour in social insect societies.

Acknowledgements

We are grateful to Dr David J.C. Fletcher for providing transcripts of his interview with Mrs J. Ingham and Mrs E.J. Walton (Mr Onions' daughters) who provided detailed personal information about Mr G.W. Onions and the photograph of G.W. Onions. Financial support was provided in part through the South African National Research Foundation (NRF) through a DSI/NRF innovation postdoctoral fellowship (grant no. 129608) to F.N.M.

Competing interests

We have no competing interests to declare.

References

1. Moritz R, Crewe R. The dark side of the hive: The evolution of the imperfect honeybee. Oxford: Oxford University Press; 2018.
2. Allsopp M, Crewe R. The Cape honey bee as a Trojan horse rather than the hordes of Genghis Khan. Am Bee J. 1993;133(2):121–123.
3. Pirk CWW, Human H, Crewe RM, Van Engelsdorp D. A survey of managed honey bee colony losses in the Republic of South Africa – 2009 to 2011. J Apic Res. 2014;53(1):35–42. https://doi.org/10.3896/IBRA.1.53.1.03
4. Onions GW. South African "fertile-worker bees". Agric J Union S Afr. 1912;3(5):720–728. https://hdl.handle.net/10520/AJA0000021_1356
5. Onions GW. South African "fertile" worker-bees. Agric J Union S Afr. 1914;7(1):44–46. https://hdl.handle.net/10520/AJA0000021_317
6. Verma S, Ruttner F. Cytological analysis of the thelytokous parthenogenesis in the Cape honeybee (*Apis mellifera capensis* Escholtz). Apidologie. 1983;14(1):41–57. https://doi.org/10.1051/apido:19830104
7. Jack RW. XXI. Parthenogenesis amongst the workers of the Cape honey-bee: Mr. G. W. Onions' experiments. Trans R Entomol Soc Lond. 1917;64(3–4):396–403. https://doi.org/10.1111/j.1365-2311.1917.tb03141.x
8. Van Warmelo D. South African fertile-worker bees and parthenogenesis. Agric J Union S Afr. 1912;3(6):786–789. https://hdl.handle.net/10520/AJA0000021_611
9. De Villiers JH. Cape bees and "Animal Intelligence". Nature. 1883;28(705):5–6. https://doi.org/10.1038/028005b0
10. Lundie A. Laying worker bees produce worker bees. S Afr Bee J. 1954;29:10–11.
11. Anderson RH. The laying worker in the Cape honeybee, *Apis mellifera capensis*. J Apic Res. 1963;2(2):85–92. https://doi.org/10.1080/00218839.1963.11100065
12. Dietemann V, Lubbe A, Crewe RM. Human factors facilitating the spread of a parasitic honey bee in South Africa. J Econ Entomol. 2006;99(1):7–13.
13. Kryger P, Dietemann V, Crewe R. Have we found a solution to the Capensis problem? S Afr Bee J. 2003;75:123–128.
14. Mumoki FN, Yusuf AA, Pirk CWW, Crewe RM. The biology of the Cape honey bee, *Apis mellifera capensis* (Hymenoptera: Apidae): A review of thelytoky and its influence on social parasitism and worker reproduction. Ann Entomol Soc Am. 2021;114(2):219–228. https://www.doi.org/10.1093/aesa/saaa056
15. Crewe RM, Velthuis HHW. False queens: A consequence of mandibular gland signals in worker honeybees. Naturwissenschaften. 1980;67(9):467–469. https://doi.org/10.1007/BF00405650
16. Hepburn HR, Crewe RM. Portrait of the Cape honeybee, *Apis mellifera capensis*. Apidologie. 1991;22(6):567–580. http://dx.doi.org/10.1051/apido:19910601

AUTHORS:
Tainã G. Loureiro[1]
Nicole du Plessis[2]
Ken Findlay[1]

AFFILIATIONS:
[1]Centre for Sustainable Oceans, Faculty of Applied Sciences, Cape Peninsula University of Technology, Cape Town, South Africa
[2]National Research Foundation – South African Environmental Observation Network (SAEON), Egagasini Node, Cape Town, South Africa

CORRESPONDENCE TO:
Tainã Loureiro

EMAIL:
loureiro.tg@gmail.com

HOW TO CITE:
Loureiro TG, Du Plessis N, Findlay K. Into the blue – The blue economy model in Operation Phakisa 'Unlocking the Ocean Economy' Programme. S Afr J Sci. 2022;118(11/12), Art. #14664. https://doi.org/10.17159/sajs.2022/14664

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
sustainability, inclusivity, sustainable development, ocean economy, ocean wealth

FUNDING:
Cape Peninsula University of Technology, DSI/NRF SARChI (UID 125455)

PUBLISHED:
30 November 2022

Into the blue – The blue economy model in Operation Phakisa 'Unlocking the Ocean Economy' Programme

Significance:

Economic and social benefits of ocean resource uses have motivated numerous nations, including South Africa, to turn to their Exclusive Economic Zones (EEZs) to advance economic development initiatives. Such initiatives result in increasing and spatially competitive pressures on ocean systems, compromising ecosystem services and market and non-market ocean benefits. It is critical to prioritise sustainable development in any ocean or blue economy advancement programmes (where the blue economy model most often parallels a terrestrial green economy, to incorporate sustainability and inclusivity pillars over and above the often GDP-centred ocean economy model). We explore multiple definitions of ocean and blue economies, discuss the importance of adopting blue economy models, and examine how the South African Operation Phakisa – Unlocking the Ocean Economy initiative presents numerous features aligned with ocean sustainable development, sustainability, and inclusivity, which strongly align it to a blue economy model.

Disentangling the ocean economy and the blue economy terminologies

Our ability to engage with peers, partners, collaborators, and stakeholders depends on our capacity to communicate. For that, adopting harmonised terms and definitions is key. However, ambiguity is identified when focusing on the concept of the blue economy. While an ocean economy is defined as the sum of market economic activities, sectors and industries supported by marine and coastal areas, their assets, goods and services[1], the term 'blue economy' has no agreed definition and has been used widely by different stakeholders in various contexts. Three main categories of use are identified: (1) as used in the lead-up to the Rio+20 UNCSD by coastal nations (particularly the Pacific Small Island Developing States) as a parallel term to 'green economy' and including sustainability and equity (or social inclusion) within the concept; (2) as used by Gunther Pauli in his book *The Blue Economy: 10 years – 100 innovations – 100 million jobs*[2] advocating innovative solutions to production efficiency and sustainability within sustainable development (and not restricted to the ocean realm); and (3) as a synonym for ocean economy, thus as any economic activity in the maritime sector, whether sustainable (and aligned with green economy principles) or not. Further confusion has resulted from the use of the terms blue and red ocean market strategies by Kim and Mauborgne[3]. Notably, beyond the three abovementioned categories, the term blue economy has been used in various contexts, including policy documents and academic publications. Such uses always refer to the advancement of ocean economies, but in many cases also include the pillars of sustainability and inclusivity (Table 1).

The broad manner in which the term 'blue economy' has been used can generate miscommunication, misunderstanding, misguided governance outcomes, and social injustice.[4,5] Although the ocean economy and blue economy terms are not well delineated, we advocate for the use of the term blue economy when referring to models that include sustainability and inclusivity (i.e. social equity) as essential elements or pillars. This approach avoids the use of blue economy as a synonym for ocean economy, supporting its use as a parallel to green economy and ocean sustainable development to define the production and growth of established and emergent ocean sectors while sharing the benefits and opportunities equitably, thus promoting social well-being and preserving ocean wealth and the health of marine and coastal ecosystems and their ecological functioning, both now and in the long term.[6] However, while there is no official definition, we suggest that authors clarify their definition a priori when using the term blue economy in papers, reports and documents, to avoid ambiguity.

Turning the tide toward blue economies

While the ocean provides humans with numerous utilities, the drivers (e.g. food security) of human activities (e.g. fisheries) result in pressures (unsustainable extraction of fisheries resources or pollution) on the ocean. Pressures are the mechanisms resulting in the change in the state of a system; in other words, they are the stressors that may harm life, health, property, or the environment.[7] Depending on the vulnerability of the receiving system, the frequency and the severity of anthropogenic pressures trigger the risk of state changes in natural systems, leading to impacts on human welfare[7] across environmental, social, and economic components. Accordingly, changes in natural capital assets may reduce (1) their opportunities for continued supply, (2) their use in production, (3) resultant volumes of produced goods and services, and (4) consequent distribution of economic benefits. Thus, there is a need to prioritise sustainable models for ocean economies, ensuring the optimisation of resource use in the present while protecting future ocean resource-use opportunities.

Globally, the importance of ocean economy development underpinned by sustainability and inclusivity is reinforced by numerous initiatives and commitments. For example, the 17 United Nations (UN) Agenda 2030 Sustainable Development Goals (SDGs) include goals and targets aimed at accelerating a nation's development while balancing social, economic, and environmental sustainability. A second example is the establishment of the UN's Decade of Ocean Science for Sustainable Development (2021–2030), which aims to motivate nations to implement activities and join efforts to generate the global ocean science needed to support and inform ocean sustainable development. Finally, it is essential to mention the constitution of the High Level Panel for a Sustainable Ocean Economy, representing 16 nations (Australia, Canada, Chile, Fiji, France, Ghana, Indonesia, Jamaica, Japan, Kenya, Mexico, Namibia, Norway, Palau, Portugal and the USA) that are committed to bridging ocean health and wealth

Table 1: Examples of the development and use of the term 'blue economy' in the policy and academic literature

Definition / use	Pillars	Reference
Policy		
'The Blue Economy conceptualises oceans as "Development Spaces" where spatial planning integrates conservation, sustainable use, oil and mineral wealth extraction, bioprospecting, sustainable energy production and marine transport' and 'will incorporate ocean values and services into economic modelling and decision-making processes'. 'At the core of the Blue Economy concept is the de-coupling of socioeconomic development from environmental degradation.'	Economic development and sustainability	UN Conference on Trade and Development[21] Building on the Rio+20 discussions
'A sustainable ocean economy [used as a synonym for blue economy] emerges when economic activity is in balance with the long-term capacity of ocean ecosystems to support this activity and remain resilient and healthy.'	Economic development and sustainability	Economist Intelligence Unit[22]
'The Blue Economy in Africa covers aquatic and marine spaces, including oceans, seas, coasts, lakes, rivers, and underground water' and 'includes recognition that the productivity of healthy freshwater and ocean ecosystems is a pathway for aquatic and maritime-based economies and can ensure that islands and other coastal countries, as well as land-locked States, benefit from their resources. It also requires an integrated, holistic and participatory approach that includes sustainable use and management of Blue Economy resources for societal progress in a diverse Africa.'	Economic development, sutainability and inclusivity	UN Economic Commission for Africa[9] Building on the Rio+20 discussions
Academic		
'The blue economy includes all industries that are dependent in some way for their development on ocean resources.'	Economic development	Laffoley et al.[23]
'…aims to tap into the estimated USD 24 trillion in potential goods and services (i.e., energy generation, mining, tourism, maritime transport, aquaculture, and capture fisheries), derived from the world's oceans, and to balance industrialization of oceans with environmental protection.'	Economic development and sustainability	Cohen et al.[24]
'…encompasses both the economic uses of the ocean and ocean resources, and the natural assets and ecosystem services that the ocean provides (sustainable ocean economy).'	Economic development and sustainability	Rayner et al.[25]

and supporting the transition to a sustainable ocean economy, among other initiatives aimed at advancing and prioritising ocean sustainable development models.

Africa has a large potential to advance blue economy models, as demonstrated by its strategic geographical position and numerous underexplored and underexploited resource-use activities that can support inclusive growth, which is critical to underpin Africa's socio-economic transformation.[8,9] Notably, the prioritisation of blue economies in Africa is imperative on various scales (i.e. continental, regional and national), as demonstrated by the African Union's (AU) Agenda 2063, which supports inclusivity and sustainability initiatives; the AU's 2050 Africa's Integrated Maritime Strategy (2050 AIMS) that highlights the critical potential of the seas to contribute towards the continent's growth and development, identifying the blue economy as the 'new frontier of African Renaissance'[10]; the publication of the African Blue Economy Strategy[8], and Implementation Plan 2021–2025[11]. A second example of the importance of blue economies in Africa is the publication of *Africa's Blue Economy: A Policy Handbook*, in 2016, by the UN Nations Economic Commission for Africa, which guides African Nations to advance the blue economy model into their national development plans, strategies, and policies.[9]

On a regional scale, the Indian Ocean Rim Association endorsed the blue economies as a cross-cutting priority area in 2014, with an emphasis on growing the blue economy in a sustainable and inclusive manner.[12] A Working Group on the Blue Economy was established in 2019, composed of government departments responsible for ocean management or development, to advance the technical capacities of Member States for the sustainable development of the ocean.

On a national scale, various African nations developed blue economy initiatives such as Go Blue in Kenya, ProAzul in Mozambique, the development of the Somali Blue Economy Centre and the publication of the 'Seychelles Blue Economy Strategic Policy Framework and Roadmap' and the report 'Towards a Blue Economy – Harnessing Namibia's Coastal Resources Sustainably'. Accordingly, the African continent is moving towards prioritising sustainable models to underpin its coastal nations' social and economic growth through ocean resource-use activities.

South Africa's Operation Phakisa

Like most coastal nations, South Africa relies on ocean resources to underpin economic development. In 2010, South Africa's ocean economy contributed approximately ZAR54 billion to the national gross domestic product (GDP) and accounted for about 316 000 jobs.[13] Moreover, the estimated potential (albeit of unreferenced methodology) contribution is up to ZAR177 billion to the GDP and over one million jobs by 2033.[14,15] South Africa has a vast ocean space, with a coastline of almost 2798 km and an Exclusive Economic Zone (EEZ) of nearly 1.5 million square kilometres, encompassing a range of oceanographic provinces and ecosystems from sub-Antarctic and cold temperate to subtropical systems.[13,16] Therefore, it is not surprising that the country is promoting ocean economy expansion to drive economic development, food and energy security, and expediting delivery of the National Development Plan (NDP).

The NDP 2030 is the national socio-economic development blueprint for eliminating poverty and reducing inequality by creating employment opportunities and redistributing and balancing access opportunities by 2030. Expediting the delivery of the NDP 2030 motivated the launch, in 2014, of the South African Operation Phakisa programme (with Phakisa meaning 'to hurry' in Sesotho). The programme encompasses various national sectors, including those related to the ocean economy encapsulated in the 'Operation Phakisa: Unlocking the Economic Potential of South Africa's Oceans' (hereafter 'Ocean Phakisa') programme. Officially launched in 2014, Ocean Phakisa was based on the analyses of the potential of nine national marine and maritime industry sectors as key priorities for advancement. This process eventually resulted in the definition of six delivery areas, namely: (1) Marine Transport and Manufacturing; (2) Offshore Oil and Gas Exploration and Production; (3) Aquaculture; (4). Marine Protection Services and Ocean Governance; (5) Small Harbours Development; and (6) Coastal and Marine Tourism.[15]

The initial phase of the programme encompassed the creation of the Oceans Economy Laboratories or 'Labs' focused on each of the delivery areas. These Labs – consisting of workshops in which key stakeholders from various sectors and institutions (public and private sectors, academia, civil society, etc.) joined efforts to identify goals and challenges, define priorities, and set plans and initiatives – resulted in detailed reports and action plans to achieve the defined targets and priorities.[13] As a result, several initiatives and key performance indicators were identified for each delivery area, together with two enabler areas: skills development and research.

Among the initiatives identified by the Labs are numerous aspects that highlight priorities aligned with a blue economy model underpinned by the pillars of economic development, sustainable exploitation and equitable share of resources and opportunities (Table 2).[14,17] A good example is the rationale of the Marine Protection Services and Governance Delivery Area, which aims at implementing 'an overarching, integrated ocean governance framework for sustainable growth of the ocean economy' and proposes the establishment of 22 new Marine Protected Areas, and the development of marine spatial planning legislation in 2016. Marine spatial planning implementation is one of the key components of this delivery area.[18] Further, under this delivery area, the South African National Oceans and Coastal Information Management System (OCIMS) platform was developed by the Department of Forestry, Fisheries & the Environment and the Department of Science & Innovation to provide information on existing ocean and coastal data services and support decision-making processes.[19]

The alignment between components of Ocean Phakisa and the blue economy model is also evidenced by the Small Harbours Development delivery area, which focuses on developing and expanding small harbours to drive economic growth through small-scale fisheries and other small coastal vessel activities, encouraging inclusive maritime development; as well as by the Aquaculture delivery area, which highlights the potential for rural development of this sector, particularly for marginalised coastal communities.[15] The inclusivity aspect of Ocean Phakisa is crucial in the context of the central pillars of the NDP so inclusive growth is one indicator metric selected to measure the programme's performance[14], and the Aquaculture Lab targets this indicator as a priority to motivate transformation in the national aquaculture industry to reduce inequality[20]. Furthermore, the implementation of Operation Phakisa is underpinned by the NDP 2030, which aims at poverty alleviation, employment generation and redistribution of access while prioritising strategies for an environmentally sustainable, climate-change resilient, low-carbon economy and just society.[15] Thus, the sustainability and inclusivity aspects of the NDP 2030 are also embedded throughout Operation Phakisa.

Conclusions

The ocean's role in supporting human well-being is only starting to be realised and recognised. Unquestionably, marine and coastal environments are essential for maintaining life on earth and supporting human development and welfare. Sustainable and inclusive strategies for ocean economy development have been highlighted and supported by numerous international, regional, and national initiatives to ensure long-term and equitably distributed use of ocean opportunities and benefits. However, an essential step for achieving the collaborative effort of developing ocean economies through sustainable models is to guarantee that stakeholders use categories, terms, and definitions in a standard way to avoid misunderstandings.

In this arena, Ocean Phakisa has numerous aspects of such a sustainable and inclusive ocean (i.e. blue) economy model that are aligned with broader national socio-economic goals, as South Africa's constitution that requires the conservation and sustainable use of the marine environment, the National Strategy for Sustainable Development which includes priorities related to the implementation of programmes

Table 2: Key initiatives or goals of each delivery area of Operation Phakisa: Unlocking the Economic Potential of South Africa's Oceans that underpin blue economy models

Delivery area	Initiative / goal
Offshore Oil and Gas Exploration	Enhancement of environmental governance capacity of the oil and gas regulator
	Exploitation of the broader research opportunities presented by offshore oil and gas exploration
Marine Protection Services and Ocean Governance	Enhancement of legislation for the *Integrated Coastal and Oceans Management Act* or Oceans Act
	Enhanced and coordinated enforcement programme for ocean protection
	Reduction in illegal and unregulated activities in the ocean space
	Reduction of pollution risks to human health and environment
	Development of a national ocean and coastal information system and extension of earth observation capacity
	Implementation of a national ocean and coastal water quality monitoring programme
	Creation of a Marine Protected Area representative network
	Development of marine spatial planning process
	Identification and protection of sensitive and unique marine habitats and species
Small Harbour Development	Establishment of new small harbours and development of coastal properties
	Redevelopment and maintenance of proclaimed small harbours
	Prioritisation of socio-economic impact through job creation, skills development and enterprise development
Coastal and Marine Tourism	To grow a world-class and sustainable coastal and marine tourism
	Publication of responsible tourism guidelines
Aquaculture	Target the inclusive growth of the aquaculture industry
Across all areas	Research, capacity building and skills development

for the sustainable management of marine ecosystems and the NDP mentioned above.

Recognising the alignment of Ocean Phakisa with blue economy models could favour South Africa's positioning in terms of national commitments to the UN's SDGs, in particular, SDG 14 (Life Below Water), SDG 1 (No Poverty), SDG 2 (Zero Hunger), SDG 8 (Decent Work and Economic Growth), SDG 13 (Climate Action), and SDG 17 (Partnerships for the Goals). This strategy would also position Ocean Phakisa in alignment with the AU priorities mentioned above and with the Indian Ocean Rim Association's blue economy regional initiatives. Additionally, framing Ocean Phakisa as a blue economy initiative could support South Africa's inclusion in organisations such as the High Level Panel for a Sustainable Ocean Economy and attract investments (i.e. ocean economy financing) focused on the transition to sustainable models.

Acknowledgements

We thank the Cape Peninsula University of Technology, the GOAP secretariat and the NRF Ocean Accounting Community of Practice for meaningful discussion and knowledge exchange. This work was supported by the Cape Peninsula University of Technology and by the South African Research Chairs Initiative through the South African National Department of Science and Innovation/National Research Foundation, by a Community of Practice grant in Ocean Accounts Framework (UID: 125455). The opinions expressed in this publication are not necessarily those of the funding bodies.

Competing interests

We have no competing interests to declare.

References

1. Organisation for Economic Co-operation and Development (OECD). The ocean economy in 2030. Paris: OECD; 2016. https://doi.org/10.1787/9789264251724-en
2. Pauli GA. The blue economy: 10 years, 100 innovations, 100 million jobs. Taos: Paradigm Publications; 2010.
3. Kim WC, Mauborgne R. Blue ocean strategy: How to create uncontested market space and make competition irrelevant. Boston, MA: Harvard Business Review Press; 2015.
4. Voyer M, Quirk G, McIlgorm A, Azmi K. Shades of blue: What do competing interpretations of the Blue Economy mean for oceans governance? J Environ Policy Plan. 2018;20(5):595–616. https://doi.org/10.1080/1523908X.2018.1473153
5. Bennet NJ, Blythe J, White CS, Campero C. Blue growth and blue justice: Ten risks and solutions for the ocean economy. Mar Policy. 2021;125, Art. #104387. https://doi.org/10.1016/j.marpol.2020.104387
6. Fenichel EP, Addicott ET, Grimsrud KM, Lange G-M, Porras I, Milligan B. Modifying national accounts for sustainable ocean development. Nat Sustain. 2020;3(11):889–895. https://doi.org/10.1038/s41893-020-0592-8
7. Elliott M, Burdon D, Atkins JP, Borja A, Cormier R, de Jonge VN, et al. "And DPSIR begat DAPSI(W)R(M)!" – A unifying framework for marine environmental management. Mar Pollut Bull. 2017;118(1):27–40. https://doi.org/10.1016/j.marpolbul.2017.03.049
8. Inter-African Bureau for Animal Resources (AU-IBAR). Africa Blue Economy Strategy. Nairobi: African Union; 2019.
9. United Nations Economic Commission for Africa (UNECA). Africa's Blue Economy: A policy handbook. Addis Ababa: UNECA; 2016.
10. African Union. 2050 Africa's integrated maritime strategy (2050 AIMS). Addis Ababa: African Union; 2014.
11. Failler P, Karani P, Gilau A, Hamukuaya H, Diop S. Africa Blue Economy Strategy Implementation Plan 2021-2025. Nairobi: AU-IBAR; 2020.
12. Indian Ocean Rim Association (IORA). Mauritius Declaration on Blue Economy [document on the Internet]. c2015 [cited 2022 Jun 09]. Available from: https://www.iora.int/media/8216/iora-mauritius-declaration-on-blue-economy.pdf
13. Potgieter T. Oceans economy, blue economy, and security: Notes on the South African potential and developments. J Indian Ocean Reg. 2018;4(1):49–70. https://doi.org/10.1080/19480881.2018.1410962
14. Operation Phakisa. Oceans economy: Summary progress report [document on the Internet]. c2019 [cited 2022 Jun 09]. Available from: https://www.dffe.gov.za/sites/default/files/docs/publications/oceans-economy-summary-progress-report-June2019.pdf
15. Operation Phakisa. Operation Phakisa oceans economy – investing and doing business with Africa's leading ocean economy [document on the Internet]. c2016 [cited 2022 Jun 09]. Available from: https://www.environment.gov.za/sites/default/files/docs/publications/southafricas_oceaneconomy__forinvestment.pdf
16. Skowno AL, Poole CJ, Raimondo DC, Sink KJ, Van Deventer H, Van Niekerk L, et al. National Biodiversity Assessment: The status of South Africa's ecosystems and biodiversity. Pretoria: South African National Biodiversity Institute; 2019. Available from: http://bgis.sanbi.org/NBA/NBA2011_metadata_formalprotectedareas.pdf%5Cnpapers2://publication/uuid/786A77C5-B11A-4F8D-B139-F3F626EBC802
17. Findlay K, Bohler-Muller N. South Africa's ocean economy and Operation Phakisa. In: Bohler-Muller N, Attri VN, editors. The blue economy handbook of the Indian Ocean region. Africa Institute of South Africa; 2018. p. 231–255. https://doi.org/10.2307/j.ctvgc60f0.17
18. Dorrington RA, Lombard AT, Bornman TG, Adams JB, Cawthra HC, Deyzel SHP, et al. Working together for our oceans: A marine spatial plan for Algoa Bay, South Africa. S Afr J Sci. 2018;114, Art. a#0247. http://dx.doi.org/10.17159/sajs.2018/a0247
19. Byrne D, Williams L, Backeberg B, Bernard S, Luck-Vogel M, McAlister B. In support of integrated coastal ocean decision-making: An introduction to the South African National Oceans and Coastal Information Management System. In: AGU Fall Meeting Abstracts; 2018 December 10–14; Washington DC, USA. p. OS21A-04.
20. Ginindza N. Operation Phakisa : Aquaculture, small harbours , marine protection services, and ocean governance [document on the Internet]. c2020 [cited 2022 Jun 09]. Available from: https://static.pmg.org.za/221124_Operation_Phakisa_Ocean_Economy.pdf
21. UN Conference on Trade and Development (UNCTAD). Blue economy concept paper. Geneva: UNCTAD; 2014.
22. The Economist Intelligence Unit (EIU). The blue economy: Growth, opportunity, and a sustainable ocean economy. London: EIU; 2015. Available from: https://www.oceanprosperityroadmap.org/wp-content/uploads/2015/05/2.-State-of-the-Blue-Economy_briefing-paper_WOS2015.pdf
23. Laffoley D, Baxter J, Lefebvre C, Sévin M-A, Simard F. Building MPA networks by 2020: IMPAC3 achievements, future challenges and next steps. Aquat Conserv. 2014;24(S2):238–245. https://doi.org/10.1002/aqc.2531
24. Cohen PJ, Allison EH, Andrew NL, Cinner J, Evans LS, Fabinyi M, et al. Securing a just space for small-scale fisheries in the blue economy. Front Mar Sci. 2019;6:1–8. https://doi.org/10.3389/fmars.2019.00171
25. Rayner R, Jolly C, Gouldman C. Ocean observing and the blue economy. Front Mar Sci. 2019;6:330. https://doi.org/10.3389/fmars.2019.00330

AUTHORS:
Charles P. Egeland[1]
Travis Rayne Pickering[2,3]
Cynthia M. Fadem[4]
Manuel Domínguez-Rodrigo[5,6]

AFFILIATIONS:
[1]Department of Anthropology, University of North Carolina at Greensboro, Greensboro, NC, USA
[2]Department of Anthropology, University of Wisconsin-Madison, Madison, WI, USA
[3]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa
[4]Gray & Pape, Inc., Providence, RI, USA
[5]Institute of Evolution in Africa, University of Alcalá de Henares, Madrid, Spain
[6]Department of Anthropology, Rice University, Houston, TX, USA

CORRESPONDENCE TO:
Charles Egeland

EMAIL:
cpegelan@uncg.edu

HOW TO CITE:
Egeland CP, Pickering TR, Fadem CM, Domínguez-Rodrigo M. Back from the dead: Another response to the contextual bases of the Rising Star 'deliberate body disposal' hypothesis. S Afr J Sci. 2022;118(11/12), Art. #13873. https://doi.org/10.17159/sajs.2022/13873

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
palaeoanthropology, taphonomy, mortality salience, *Homo naledi*, Rising Star Cave

PUBLISHED:
30 November 2022

Back from the dead: Another response to the contextual bases of the Rising Star 'deliberate body disposal' hypothesis

Significance:

The hypothesis that >1500 Middle Pleistocene hominin bones represent the remains of complete corpses deposited deliberately in Rising Star Cave by conspecifics is provocative. This is because intentional handling of dead bodies might imply these hominins had developed a uniquely human sense of mortality salience >235 000 years ago. We assess the contextual bases of this hypothesis and find they do not, in fact, provide its unequivocal support. In sum, critical assessment of relevant geological and taphonomic data disallows falsification of the null hypothesis that the assemblage formed as the result of a non-anthropogenic process(es). Because so, the 'deliberate body disposal' hypothesis remains unsupported.

The large assemblage of hominin fossils from Rising Star Cave (RSC) (South Africa) is one of the most remarkable palaeoanthropological finds ever made. Analyses of its contents expand our understanding of the taxonomy, functional morphology, and tempo and mode of human evolution. Perhaps the most intriguing aspect of the RSC fossils is, however, what they might be able to tell us about the behaviour of the hominin individuals from which they derive. Most dramatically, it is argued that the RSC bones are from the corpses of hominins that were placed deliberately in the cave by conspecifics. Fossil-bearing areas in addition to the original Dinaledi Chamber have recently been recorded in the RSC system, including the U.W. 110 locality that has yielded 34 craniodental specimens of a single hominin juvenile.[1] Allusion to the stratigraphic context of these newly described fossils – i.e. that it appears 'similar to the U.W. 102b area of the Lesedi Chamber, in which cranial fragments and teeth from a single, immature individual were found on the surface and in the shallow, sub-surface contexts in sediments that rested on a horizontal chert shelf approximately 80cm above the cave floor'[1(p.10)] – has reignited scientific and popular discussion of the extraordinary claim that it is 'likely that some [intentional] hominin agency was involved in the deposition'[1(p.13)] of the RSC hominin materials. We appreciate the challenge RSC poses to researchers in terms of its difficult working conditions and complex stratigraphy. Videos of the cave system not only reinforce this appreciation but, in combination with published descriptions of its stratigraphy, geomorphology, and fossil assemblages, also afford us this opportunity to comment on two salient contextual underpinnings of what Dirks et al.[2(p.29)] refer to as the 'deliberate body disposal' hypothesis.

1. Entry into the Dinaledi sub-system today requires travel up the Dragon's Back (DB) and then descent through a narrow 12-m vertical fissure termed the 'Chute'. Elliott et al.[3(p.16)] reiterate the assertion from previous publications that 'the Chute was the only viable access point during the time that the hominin material accumulated, and thereafter'. The DB is a large dolomite block that, after its detachment from the cave ceiling at some point in antiquity, posed a significant impediment to the access of sediments and fauna into the Dinaledi sub-system, including especially hominins that might have been transporting deceased conspecifics. The timing of the DB collapse is thus of considerable significance for the 'deliberate body disposal' hypothesis. Robbins et al.[4(p.19)] date this event provisionally to sometime between 290 and 225 kiloannum (ka), which coincides broadly 'with the period when [hominins] most likely interacted with the Dinaledi sub-system'. It seems, then, that prior to this time range, access into the Dinaledi sub-system would not necessarily have required a journey up and over the DB and then down the Chute but perhaps, instead, a more manageable 5-m ascent up a dolomite sill followed by a short span *under* the still-attached DB block. This scenario helps make sense of the laminated orange-red muds found throughout the RSC system. Although some of these deposits may have formed through vadose water drip, most are interpreted as slackwater sediments that resulted from suspension settling of hydrologically transported clay and silt.[5] Erosional remnants of these muds occur throughout the cave system, including some several metres above the current floor in the Dinaledi Chamber itself. The arrival of these formerly extensive deposits must have involved water that originated outside the Dinaledi sub-system and travelled through the DB Chamber some 350 ka[4(p.18)], and, thus, before the deposition of the hominin fossils. Fluvial activity of some kind also must be implicated in these sediments' eventual erosion and dispersal deeper into the cave system via floor drains during and after the deposition of the hominin fossils. This not only highlights accessibility to the Dinaledi sub-system but also demands consideration of non-anthropogenic (i.e. hydraulic) mechanisms of bone deposition in the cave. Of course, such accessibility would also apply to hominins as they (potentially) moved into and out of the cave's chambers. However – and importantly – the recovery of baboon remains in a fissure just off the Chaos Chamber[3] demonstrates that the mere presence of large-bodied primates deep within the cave system need not necessarily require deliberate disposal by conspecifics.

2. Three key claims stand out to us about the taphonomic history of the hominin fossils from the Dinaledi Chamber[2]: (1) complete corpses were deposited within the chamber; (2) there is a lack of damage created by mammals on the hominin fossils; and (3) the remains were never exposed on the landscape outside of the cave. We have already scrutinised the first two claims elsewhere.[6] In summary: (1) the published RSC hominin skeletal part frequencies indicate either that complete corpses did not enter the chamber, or that, if they did, they then experienced some level of disturbance after their original deposition within the chamber; and (2) the poorly preserved cortices of the fossils may have obscured evidence of carnivore tooth marks and other taphonomic damage. The legitimacy of these concerns rests on how closely the frequencies of skeletal

parts from the 2013–2014 excavated assemblage match that of the portion of the assemblage that remains in the chamber as well as on the results of a comprehensive microscopic analysis of the cortical surfaces of all the fossils. The third claim relies, in part, on patterns of sub-aerial weathering damage to bone cortices. Linear cracks consistent with Behrensmeyer's[7] Stage 1 are common on the hominin fossils. Dirks et al.[2(p.33)] acknowledge that these standards – which, after all, were constructed for bones exposed aboveground – are not completely applicable to subsurface environments. They do assert, however, that because the bones do not exhibit damage indicative of extended surface exposure, the observed weathering must be due to the swelling and shrinkage of bone exposed to wet-dry cycles within the recesses of the cave. Pokines et al.[8] do in fact demonstrate through actualistic work that moisture-induced swelling and shrinkage can produce linear cracks similar to those observed on the hominin bones. However, they also note[8(p.438)] that it is 'unknown if fluctuations in interior karst feature humidity can bring about sufficient wet-dry cycles to exposed surface bone to cause weathering changes, so this topic of research needs exploration in field and additional laboratory settings'. To our knowledge, actualistic work has not confirmed that sub-aerial weathering damage to macromammal bones can occur deep within a cave system. Even if this were the case, environmental conditions within RSC are relatively constant today and were likely 'stable and dry for at least the last ca. 300 ka'[5(p.914)], so it remains an open question whether the RSC bones experienced wet-dry cycles of sufficient amplitude to create the observed weathering damage.

In light of these concerns, we continue to question the plausibility of the RSC 'deliberate body disposal' hypothesis. We agree that the RSC fauna is exceptional for its extremely high proportion of hominin fossils. But careful consideration of hominin skeletal part representation, acknowledgement that poorly preserved bone cortices have biased the identification of possible surface damage, and, as discussed here, appreciation of the potentials that the RSC fauna is hydraulically derived and/or is wholly or partially autochthonous, all mean that its formation is still quite uncertain.

References

1. Brophy JK, Elliott MC, De Ruiter DJ, Bolter DR, Churchill SE, Walker CS, et al. Immature hominin craniodental remains from a new locality in the Rising Star Cave System, South Africa. PaleoAnthropology. 2021;2021:1–14.
2. 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, e09561. https://doi.org/10.7554/eLife.09561
3. Elliott MC, Makhubela TV, Brophy JK, Churchill SE, Peixotto B, Feuerriegel EM, et al. Expanded explorations of the Dinaledi Subsystem, Rising Star Cave System, South Africa. PaleoAnthropology. 2021;2021:15–22.
4. Robbins JL, Dirks PHGM, Roberts EM, Kramers JD, Makhubela TV, Hilbert-Wolf HL, et al. Providing context to the *Homo naledi* fossils: Constraints from flowstones on the age of sediment deposits in Rising Star Cave, South Africa. Chem Geol. 2021;567, Art. #120108. https://doi.org/10.1016/j.chemgeo.2021.120108
5. Wiersma JP, Roberts EM, Dirks PHGM. Formation of mud clast breccias and the process of sedimentary autobrecciation in the hominin-bearing (*Homo naledi*) Rising Star Cave system, South Africa. Sedimentology. 2020;67(2):897–919. https://doi.org/10.1111/sed.12666
6. Egeland CP, Domínguez-Rodrigo M, Pickering TR, Menter CG, Heaton JL. Hominin skeletal part abundances and claims of deliberate disposal of corpses in the Middle Pleistocene. Proc Natl Acad Sci USA. 2018;115:4601–4606. https://doi.org/10.1073/pnas.1718678115
7. Behrensmeyer AK. Taphonomic and ecologic information from bone weathering. Paleobiology. 1978;4:150–162. https://doi.org/10.1017/S0094837300005820
8. Pokines JT, Faillace K, Berger J, Pirtle D, Sharpe M, Curtis A, et al. The effects of repeated wet-dry cycles as a component of bone weathering. J Archaeol Sci: Reports. 2018;17:433–441. https://doi.org/10.1016/j.jasrep.2017.11.025

AUTHOR:
Paul Pettitt[1]

AFFILIATION:
[1]Department of Archaeology, Durham University, Durham, United Kingdom

CORRESPONDENCE TO:
Paul Pettitt

EMAIL:
paul.pettitt@durham.ac.uk

HOW TO CITE:
Pettitt P. Did *Homo naledi* dispose of their dead in the Rising Star Cave system? S Afr J Sci. 2022;118(11/12), Art. #15140. https://doi.org/10.17159/sajs.2022/15140

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
burial, funerary caching, caves, *Homo naledi*, taphonomy

PUBLISHED:
30 November 2022

Did *Homo naledi* dispose of their dead in the Rising Star Cave system?

Significance:

Human treatment of the dead is one of the most visible and important aspects of our behavioural evolution. Until recently, the deliberate movement of corpses to specific places in the landscape and their deposition there was thought to emerge very late in human evolution, perhaps with the advent of burial by *Homo sapiens* and the Neanderthals. The remains of *Homo naledi* in South Africa's Rising Star Cave system potentially revolutionalises that belief: did a small-bodied, small-brained hominin drag parts of corpses into the depths of the cave, and if so, what does this reveal about their cognition? How convincing is the case?

The remarkable discovery of the remains of at least 15 individuals of the small-bodied and small-brained hominin *Homo naledi* in the Cradle of Humankind's Rising Star Cave system adds significantly to a growing picture of the speciose and complex nature of evolution even within the later genus *Homo.* And in a world in which genetics have revealed a large part of this complexity, the discovery is a welcome reminder that palaeoanthropology is still predicated on exploration and the excavation of physical materials. As ever, the Cradle of Humankind provides evidence of the primacy of Africa for hominin evolution, and in particular the importance of the southern parts of the continent in that story. While the rich *Homo naledi* hypodigm offers much for the study of anatomical and biological evolution, I shall restrict my comments to the hypothesis that the remains were introduced into the cave as a form of 'deliberate body disposal' rather than by any natural agent of accumulation, as considered by Dirks et al.[1] The Rising Star team have been balanced in their consideration of competing hypotheses for the introduction of the *Homo naledi* material into the cave, including an initial consideration of funerary caching given that there was a lack of any animal remains other than microfauna, a partial juvenile baboon and partial owl in a cave yielding >1500 remains of *Homo naledi*, in addition to the lack of clear indicators of natural causes of the accumulations such as transport by water, catastrophe or carnivore.

It remains possible that the Rising Star Cave represents an early expression of funerary activity by the genus *Homo.* As it stands, I am broadly in support of the team's working conclusions in this light, and although I will suggest that there are one or two factors that still need to be elucidated, I want to consider issues relating to the debate, particularly to our paradigmatic biases and how they influence our reception of the funerary caching hypothesis, before considering the argument as it stands today in the light of the cave system and the hominin material.

Paradigmatic reservations: Is funerary caching plausible?

While a barn owl and a baboon may have got lost in the cave, and died there[2], it has been difficult to identify a 'natural' cause such as this for the deposition of the *Homo naledi* remains in and around the Dinaledi sub-system. Although Dirks et al.[1] acknowledged that 'mass mortality of groups of hominins within the Dinaledi Chamber, due to a death-trap scenario, is possible', this was clearly not a singular event; why would at least 15 individuals continue to explore the depths of a cave, only to get lost and die, repeatedly over a period of time? If, by contrast, it can be demonstrated that this reflects deliberate behaviour, it has potential implications for our understanding of human cognitive and behavioural evolution. No surprise, then, that the issue has been subject to debate. One of the several hypotheses the Rising Star team consider is that bodies – or parts of them – were deposited deliberately as a form of mortuary activity – 'funerary caching' as I named it[3]. But as Randolph-Quinney[4] noted, this notion that a small-brained and relatively primitive hominin could repeatedly dispose of the dead in a deep cave 'is bound to meet with resistance'. One assumption we have all made is that if the body parts were deliberately placed in the cave, then the agent responsible for this was *Homo naledi,* rather than another, perhaps larger-brained hominin. Leaving that speculation aside, Dirks et al.[1(p.152)] ask, 'should we be surprised at the idea of a small-brained hominin species caching bodies in an inaccessible place?'. My answer to this is resoundingly no – we shouldn't. I have no objection to the notion per se: while the transport of bodies deep into a cave system required a 'non-trivial expenditure of effort', I see no reason why the exploration of an underground system by a small-brained human species should require a 'surprisingly high degree of knowledge' that needed to be 'passed on from generation to generation' as one critic suggested.[5(p.146)] As Dirks et al.[1] noted, non-hominin primates exhibit considerable variability in their reactions to and treatment of the corpses of their conspecifics and a glance beyond the primate world reveals complexity in mortuary behaviour widely in the animal world[6]. As methodologies for investigating primate behaviour towards the dead emerge[7], we may well lose our surprise that funerary caching in the genus *Homo* occurred, from time to time, among several taxa.

From the 1980s, hominin funerary behaviour – at least burial – has been set up reverentially on the altar of cognitive and behavioural modernity, worshipped as a sacred trait on the checklist of 'behavioural modernity' and monotheistically assigned, like art, exclusively to *Homo sapiens.* I suspect this has made palaeoanthropologists averse to the notion that 'pre-modern' humans had funerary practices, as many are to the notion that Neanderthals created art. True, hypotheses need testing, but an incorrect and *Homo*-centric paradigm that exaggerates the cognitive sophistication of dealing physically with dead conspecifics is of little help. Termites remove their dead from nests and cover them with sediments, and although such caching and burial behaviours (as a palaeoanthropologist might refer to them) are chemically induced and relate to homeostasis, the practice reveals that such behaviour is far from exceptional. Critics of the notion forget that bodies may be deposited in deep caves without any sophisticated cognitive rationale behind the behaviour: the problem perhaps should be seen not so much as recognising the act of curation and deposition of the corpses of conspecifics in a specific place, but how to identify

the origin of complex ideas which eventually came to be associated with such behaviour. Let's forget the latter for now.

How do we know what funerary – at present, in fact, any – behaviour *Homo naledi* was capable of? The available endocast from the Dinaledi Chamber has been taken to suggest that the taxon shared the ability for 'serialised communication, planning, and complex action', an 'increased display of prosocial emotions' and the possession of mental 'sequences that underlie tool production' that distinguished the genus *Homo* from the australopithecines. This is despite the retention of a brain intermediate in size between *Australopithecus sediba* and *Homo erectus*.[8] If this is the case, then perhaps in the case of early funerary behaviour it was not brain size that counted. One thing that such a practice would require is light. Why has no evidence of artificial light (in the form of charcoal fragments in Units 2 or 3, or torch wipes on the cave's walls) been recovered? Despite the evidence of post-depositional reworking, one might expect fragments of charcoal to remain in Unit 3 at least. Might any of the stalactites contain evidence of soot?

We need to be open to honest considerations of such hypotheses, at least when natural causes can be ruled out. But to what extent can these natural causes be excluded?

Was there another opening at the time of deposition of the hominin remains?

The Rising Star Cave system inclines downwards from the current entrance an average of 17° and the Dinaledi Chamber is 30 m underground and 80 m in a straight line from the current opening, at least 10 m below the level of the cave's palaeo-water table.[9] The 'extremely numerous and concentrated' *Homo naledi* remains in the Chamber derive from 'largely unconsolidated mud-rich sediments deep inside the cave away from any obvious cave opening which suggests that a special confluence of circumstances contributed to their accumulation'[9(p.3-4)]. These circumstances certainly included 'several cycles of sediment-flowstone fill and removal/dissolution as the level of the water table in the cave changed repeatedly'[9(p.4)]. Angular mud clasts interpreted as reflecting 'minimal transport and low-energy processes in the cave chamber' accumulated as a debris cone that 'largely developed in the *absence of either sustained flowing or standing water*'[9(p.6)] (my emphasis) and (in the case of Unit 2, which contained several fragmentary hominin remains) which 'gradually slumped into the Dinaledi Chamber'[9(p.11)]. The overlying Unit 3, which contains the bulk of the hominin material, derives from the reworking of Unit 2 mud clasts into a brown muddy matrix which accumulated across the chamber and penetrated some but not all side passages[9(p.12)]. Mineralogical comparison of the Unit 2 and 3 sediments with those of the upslope 'Dragon's Back' Chamber – through which on current evidence the material would have had to move if it did derive from downslope movement – reveals that, while the deposits of the Dragon's Back Chamber are consistent with deriving from downslope movement of allochthonous material, the Dinaledi Chamber 'was an isolated sedimentary environment at the time of deposition of Unit 3, with *no or very limited transfer of sediment between the two chambers*'[9(p.14)] (my emphasis). The weathering states of the bones are uniform and 'are consistent with the effects of sub-aerial and sub-surface processes in a *periodically wet or water-saturated* dark depositional environment that experienced stable temperatures[9](p.22) (my emphasis). Despite this, the sediments of the Dinaledi Chamber are seen as having been formed in relatively dry autochthonous conditions, apparently eliminating fluvial transport as the cause of deposition and hence it is 'highly unlikely that the fossils were washed into the cave'[10].

While 'the Dinaledi Chamber was an isolated sedimentary environment'[1(p.150)], isolated at least from the upslope Dragon's Back Chamber, the question remains as to whether another entrance, now sealed, was responsible for the introduction of sediments and hominin remains, as Val[5] hypothesised. The lack of smoothing, rounding, abrasions and impact marks on the hominin remains may 'preclude transportation by water as a major taphonomic factor associated with the delivery of skeletal elements into, or within the cave system'[1(p.23)], but can one confidently eliminate seasonal, periodic *low-energy* water transport as responsible, akin to the 'creep towards floor drains…[which] removed sediment from the chamber and caused fossils to move' in the Dinaledi Chamber[1(p.150)]? It is otherwise difficult to explain the 'high-degree of bone breakage and fragmentation' of the hominin material through 'post-depositional sediment movement within the chamber as Units 2 and 3 are reworked'[1(p.23)]. If this explains the post-depositional movement and weathering of hominin remains *within* the Dinaledi Chamber, it could explain their *introduction into* the chamber from without. Dirks et al.[1(p.150)] suggest that it would be odd if such an external link existed but that the remains of other animals were not introduced into the cave in a way similar to the hominins. Was another route into the Dinaledi Chamber once open? The wider exploration and survey of the system by Elliott et al.[11] reveals a complex picture which renders this a possibility.

This question was not addressed in an earlier publication of laser scanning and photogrammetric recording of the entire Dinaledi Chamber.[12] What emerged from that was a somewhat different vertical relationship between the Dinaledi and Dragon's Back Chambers than previously published schematic sections suggested, wherein the former is directly below the latter, rather than aside it and linked only by The Chute at ceiling level in the Dragon's Back[12] (see their Figure 3). The subsequent exploration of the wider system[11] significantly expanded the known extent of the interlinked passages and chambers, revealing other routes of possible ingress into the Dinaledi Chamber, several of which contained *Homo naledi* fossil deposits, including the U.W. 108 locality[13]. While Elliott et al.[11(p.16)] state that 'geological and speleological investigations both on the surface and underground have failed to find another entry into the area' (aside from The Chute), their discussion refers only to an aboveground survey. To my knowledge, a vertical section of the current extent of the system has not been published, but it is clear, at least from the plans, that the system is a complex set of parallel and perpendicular vertical fissures in the dolostone, each apparently with distinct sedimentary fills. The nature of these – including collapsed dolomite blocks – 'makes it difficult to determine how deep the fissures penetrate'[11(p.20)]. Although Elliott et al. refer to the fact that all but one of the newly mapped fissures are non-navigable (less than 25 cm in width), this would presumably be no obstacle while they were clear of fill, given that the team have been capable of descending The Chute, which narrows to 20 cm in places[12]. The possibility would presumably be easier still for the relatively smaller bodies of *Homo naledi* as for a baboon and owl.

Three localities deeper into the system than the Dinaledi Chamber (from the perspective of its current entrance) contained fossils pertinent to the discussion: the six skeletal elements of the baboon (U.W. 109), cranio-dental fragments of a juvenile *Homo naledi* individual (U.W. 110), and 33 elements including fragmentary long bones again consistent with *Homo naledi* (U.W. 111), all 'extremely difficult and remote localities' suggesting 'potentially different depositional events and processes from the…larger chambers'[11(p.21)], but these are all downslope from the Dinaledi and other chambers: would not a parsimonious interpretation be that *low energy* downslope mud movement during periods in which the cave system was *relatively* open (i.e. with no flowstone formation) account for the odd head or limb element penetrating into the system from a still not fully mapped fissure?

As flowstone activity – which continues today – has remodelled the cave interior, Dirks et al.[9(p.27)] acknowledge that an easier or more direct access to the Dinaledi Chamber may have existed in the past, although the sedimentology strongly suggests that the hominin-bearing matrix accumulated below the modern access point, but this rules out only the Dragon's Back route. As flowstone formation in the system seems to have occurred in discrete phases during the Middle Pleistocene[14], we might infer that the relative openness and closure of the chamber varied over this time: can another entrance (or more) be confidently ruled out?

The human remains

Much has been made of the *articulated* nature of some of the hominin remains: these include a lower limb of a child, a hand of an adult and two other partial hand articulations, an ankle, and at least four partial foot articulations.[1] As Val[5] argued, this is a very low proportion among a large sample of highly disturbed hominin remains, in fact lower

than the degree of articulation in a sample of *Australopithecus sediba* (minimum number of individuals =2) and could be explained by natural mummification, a process that Dirks et al.[1(p.151)] acknowledge they cannot rule out. Not that this need contradict the notion of funerary caching: in this light the curation of the naturally mummified remains of infants by their chimpanzee mothers reminds us that such remains can be carried around deliberately.[15] But these are articulated *body parts,* not bodies, even accounting for the post-depositional disturbance.

Remaining questions

The impressive work undertaken in the system to date continues to astound, and the team is narrowing down the possible interpretation for the introduction of the *Homo naledi* material into the cave. The funerary hypothesis certainly cannot be ruled out, and if anything it seems to me that the team nudge closer to firming up a justification for this. I'm *nearly* convinced, but not quite. Three questions for the team come to mind:

- Can one definitively rule out a different and as-yet unmapped entrance into the Dinaledi Chamber (whether or not it facilitated natural deposition of the hominin remains)?
- Is there any evidence of artificial lighting in the cave system, e.g. torch wipes, charcoal fragments or soot trapped in carbonate deposits?
- Is there evidence that it was dead *bodies,* rather than *body parts* that were carried into the chamber?

References

1. Dirks PHGM, Berger LR, Hawks J, Randolph-Quinney PF, Backwell LR, Roberts EM. Comment on 'Deliberate body disposal by hominins in the Dinaledi Chamber, Cradle of Humankind, South Africa?'. J Hum Evol. 2016;96:149–153. https://doi.org/10.1016/j.jhevol.2016.04.007
2. Kruger A, Badenhorst S. Remains of a barn owl (*Tyto alba*) from the Dinaledi Chamber, Rising Star Cave, South Africa. S Afr J Sci. 2018;114(11/12), Art. #5152. https://doi.org/10.17159/sajs.2018/5152
3. Pettitt P. The palaeolithic origins of human burial. London: Routledge; 2011. https://doi.org/10.4324/9780203813300
4. Randolph-Quinney PS. A new star rising: Biology and mortuary behaviour of *Homo naledi*. S Afr J Sci. 2015; 111(9/10), Art. #a0122. https://doi.org/10.17159/sajs.2015/a0122
5. Val A. Deliberate body disposal by hominins in the Dinaledi Chamber, Cradle of Humankind, South Africa? J Hum Evol. 2016;96:145–148. https://doi.org/10.1016/j.jhevol.2016.02.004
6. Anderson J, Biro D, Pettitt P. Evolutionary thanatology. Philos Trans R Soc Lond B. 2018;373(1754), Art. #20170262. https://doi.org/10.1098/rstb.2017.0262
7. Pettitt PB, Anderson J. Primate thanatology and hominoid mortuary archaeology. Primates. 2019;61:9–19. https://doi.org/10.1007/s10329-019-00769-2
8. Holloway RL, Hurst SD, Garvin HM, Hawks J. Endocast morphology of *Homo naledi* from the Dinaledi Chamber, South Africa. Proc Natl Acad Sci USA. 2018;115:5738–5743. https://doi.org/10.1073/pnas.1720842115
9. 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, e09561. https://doi.org/10.7554/eLife.09561.021
10. Wiersma JP, Roberts EM, Dirks PHGM. Formation of mud clast breccias and the process of sedimentary autobrecciation in the hominin-bearing (*Homo naledi*) Rising Star Cave system, South Africa. Sedimentology. 2020;67(2):897–919. https://doi.org/10.1111/sed.12666
11. Elliott MC, Makhubela TV, Brophy JK, Churchill SE, Peixotto B, Feuerriegel EM, et al. Expanded explorations of the Dinaledi Subsystem, Rising Star Cave System, South Africa. PaleoAnthropology. 2021;1:15–22.
12. Kruger A, Randolph-Quinney P, Elliott M. Multimodal spatial mapping and visualisation of Dinaledi Chamber and Rising Star Cave. S Afr J Sci. 2016;112(5/6), Art. #2016-0032. https://doi.org/10.17159/sajs.2016/20160032
13. Bolter DR, Elliott MC, Hawks J, Berger LR. Immature remains and the first partial skeleton of a juvenile *Homo naledi*, a late middle pleistocene hominin from South Africa. PLoS ONE. 2020;15(4), e0230440. https://doi.org/10.1371/journal.pone.0230440
14. Dirks PHG, Roberts EM, Hilbert-Wolf H, Kramers JD, Hawks J, Dosseto A, et al. The age of *Homo naledi* and associated sediments in the Rising Star Cave, South Africa. eLife. 2017;6, e24231. https://doi.org/10.7554/eLife.24231
15. Biro D, Humle T, Koops K, Sousa C, Hayashi M, Matsuzawa T. Chimpanzee mothers at Bossou, Guinea carry the mummified remains of their dead infants. Curr Biol. 2010;20:R351–R352. https://doi.org/10.1016/j.cub.2010.02.031

AUTHORS:
Stephanie Burton[1]
Francis Faller[2]
Andrew Kaniki[3]
Andrew Leitch[4]
Isaac Ntshoe[5]

AFFILIATIONS:
[1]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa
[2]School of Education, University of the Witwatersrand, Johannesburg, South Africa
[3]Research Management and Higher Education Consultant, Johannesburg, South Africa
[4]Department of Physics, Nelson Mandela University, Gqeberha, South Africa
[5]Unit of Research in Scholarship of Teaching and Learning (RSoTL), Central University of Technology, Bloemfontein, South Africa

CORRESPONDENCE TO:
Stephanie Burton

EMAIL:
Stephanie.burton@up.ac.za

HOW TO CITE:
Burton S, Faller F, Kaniki A, Leitch A, Ntshoe I. Are South African doctoral qualifications educating the thinkers we need? S Afr J Sci. 2022;118(11/12), Art. #14314. https://doi.org/10.17159/sajs.2022/14314

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
doctoral qualification, doctoral training, South African national report, PhD, postgraduate education

PUBLISHED:
30 November 2022

Are South African doctoral qualifications educating the thinkers we need?

Significance:

The recently completed national review of the doctoral qualifications offered by South African higher education institutions has provided important insights into the national landscape of doctoral education, and raised many questions. One key question is whether our doctoral qualifications educate our students to be the broad and critical thinkers needed to address current and future scientific and societal challenges. In the South African higher education context, we must ask ourselves whether we are providing the academic and intellectual depth required to enable our doctoral graduates to achieve the graduate attributes that we express as our national aspirations, and we need to consider new approaches to doctoral education.

Introduction

The *Doctoral Degrees National Report*[1] of the Council on Higher Education (CHE) was published in June 2022, following an intensive process of reporting and review. All South African universities and private higher education institutions which offer doctoral degree qualifications were required to undertake a Self-Evaluation Review, based on the National Standard for Doctoral Qualifications established by the CHE in 2018.[2] Individual Review Panels then reviewed the Self-Evaluation Reviews, conducted site visits to the institutions, and provided feedback to the CHE. The comprehensive National Report was compiled from the Review Panel reports and additional information provided by the CHE.

This is the first time a national review of an entire qualification type has been conducted (although previous national reviews have been conducted on the disciplinary programmes MBA, Education, Social Work, and LLB). The outcomes of this review are of tremendous significance for higher education institutions and broader society. In this Commentary, we highlight some issues that emerged from the review. The Commentary raises important questions about the extent to which we are successfully preparing our doctoral graduates as critical thinkers, equipped for future roles, and contributing effectively to addressing local, national and global scientific and societal challenges, as they are identified in the graduate attributes set out in the Qualification Standard, and we suggest some new approaches to doctoral education for consideration.

The purpose of a doctoral degree

The national Qualification Standard[2] states that the purpose of the doctoral qualification is to 'develop the highest level of holistic and systematic understanding of scholarship in, and stewardship of, a field of study through an original contribution that advances the frontiers of knowledge'. These expectations are the essence of doctoral study, emphasising high-level intellectual thinking, and distinguishing it from master's-level training programmes.

Further to the statement of purpose, the Standard sets out graduate attributes that would demonstrate the achievement, by the doctoral student, of required levels of knowledge and proficiency. These attributes resonate with global trends, in specifying outcomes that all doctoral graduates should have attained on completion of the qualification, in preparation for ongoing contribution to a research community. We note that the national review of doctoral qualifications was not intended to include a review of the Standard itself, but rather, to use it as a benchmark. In a separate article[3], we have explored in more depth the concept of graduate attributes, and the effectiveness of institutional practices in developing an understanding, among academics and students, of their significance.

The attributes are broadly categorised as 'Knowledge and Skills Attributes'. Knowledge Attributes include:

> broad, well-informed, and current knowledge of the field or discipline; insight into the interconnectedness of the topic of research with other cognate fields; ethical awareness; originality; and capacity for reflection, critical thinking and problem solving.

These relate to the expected original contribution, and its integration into existing literature and academic debate. The doctoral graduate should demonstrate:

> specialised, in-depth knowledge within a specific area of research, and awareness of the significance of their work in the field of research. The notion of originality itself requires the doctoral graduate to be a 'well-informed expert'.

Skills Attributes relate to the selection and application of appropriate research approaches and methods to answer research questions, the ability to work independently, to substantiate and defend findings and conclusions, to reflect critically on the research process, and to demonstrate critical and analytical thinking, and intellectual competence, in problem solving in diverse contexts. Furthermore, graduates should demonstrate:

> *an advanced level of communicative competence, through capacity for extended, sustained and rigorous academic writing, including relevant digital literacy skills appropriate for doctoral research, and the ability to relate individual research with reference to, and critical analysis of, associated research produced by scholars in the relevant intellectual and knowledge domain.*[2]

Such skills should encompass communication with both expert and non-expert audiences.

It is expected (by the CHE in the first instance) that institutions should take this expression of the purpose of the doctoral degree into account in planning doctoral programmes, and make this the foundation for the provision of doctoral education. The graduate attributes express the expectation that our doctoral graduates will have developed into broad and critical thinkers, implying that our doctoral programmes should be structured and implemented in ways that support such development.

The National Report considered several aspects of doctoral education that would influence the intellectual development of doctoral students as well-informed experts and critical thinkers. Some key findings and questions are summarised here.

Awareness of graduate attributes

A first, and critical, aspect of directing the structuring and implementation of doctoral programmes is clear awareness of the graduate attributes, by students, supervisors, academics and university leaders. The question we must ask is: are academic role players paying enough attention to the desired attributes and their attainment by doctoral graduates?

The Report demonstrated a surprising lack of awareness, or depth of understanding, among many academic role players, of the attributes that a doctoral graduate should attain, and especially those specific to a doctoral qualification, as opposed to generic graduate attributes. Of even greater concern, supervisors and doctoral students, in many cases, were uninformed about the Standard, and therefore had paid little or no focused attention to the graduate attributes. It was clear that students, particularly, had insufficient awareness or understanding of the developmental progression required, in realising these attributes.

Preparedness of candidates

Many factors influence the successful achievement of the doctoral qualification. Key initial steps include the process of selection of the candidate, the progression from admission to acceptance of the research proposal, and initiation of the research project. A common comment in the reviews was that our students are not well-prepared for doctoral study, and we must ask whether our institutions are providing appropriate support, to address this challenge.

The Review found that there is wide variation, among institutions, in the level of preparedness expected of doctoral candidates, at the time of their selection and admission. There is also wide variation in the provision of pre-registration measures used to assess the readiness of a candidate to engage with learning at doctoral level, and programmes to assist them in preparing for the doctoral study. There are significant variations in processes for supporting the student in the preparation and acceptance of the research proposal, a key milestone at the start of any doctoral study. In some instances, doctoral candidates are registered (or pre-registered) and provided with strong academic support in preparing their research plans; in others, prospective candidates are expected to write and submit a proposal with minimal or no support, before they are admitted. Such inconsistencies inevitably mean that many students are intellectually ill-prepared to begin doctoral studies, which may then hinder their academic development.

Supervisory support

Throughout a doctoral study, the supervisory support provided to a doctoral candidate is a pre-eminent, and critical factor. The progress of the doctoral candidate is influenced strongly by the way in which the supervisory process is conducted, and the extent to which the student's intellectual development is prioritised in supervisory guidance and supporting interactions.

The Report highlights serious concerns expressed by institutions, regarding insufficient numbers of adequately qualified supervisors in the sector, and a national emphasis on increasing numbers of doctoral graduates. These, combined, create pressure on individuals and institutions, compromising their capacity to provide high-quality supervision.

This pressure can potentially lead to conflict for supervisors, in terms of prioritising either the completion of an acceptable thesis or the development of the student, which may take more time than the national expectation of 3 years. In addition, some students may see a doctoral degree simply as a way to gain better employment in due course, as opposed to being motivated towards developing themselves intellectually. Thus, the motivation of the student for undertaking a doctoral study should be part of the admission process, and awareness of this should be part of supervisor induction.

There is clearly a need for additional supervisory capacity across the national system. While it is recognised that programmes for training supervisors are in place in some institutions, we should be asking why these are generally not mandatory, and why there is usually no certification of the training. This leads us to question how we ensure the quality, and depth, of the supervision provided to our doctoral students. In the rapidly changing national and global contexts, and with current inter-disciplinary approaches, the attitudes of institutions and supervisors should be focused on innovation and renewal, rather than remediation. Furthermore, there is much advantage in ensuring the transfer of experience from the old to the new, at institutional as well as individual level.

In all South African universities, the one-on-one supervisor-student model is the most common (despite international trends towards alternative supervision models). The student-supervisor relationship is therefore the primary source of guidance for the student, and the intellectual development of the doctoral candidate is highly dependent on the availability of a supervisor with relevant qualifications, wisdom, and experience. The National Report makes recommendations for adoption of alternative models such as cohort supervision or supervisory panels, and some newer, alternative approaches are discussed later in this article, which would promote greater depth and diversity in supervision practices.

Where a student does have more than one supervisor, the addition of co-supervision may provide a constructive route to stronger support, but there were also reports of conflicting views or approaches between supervisors, creating difficulties and negatively influencing students' progress. This influence is even more evident for a doctoral student undertaking interdisciplinary studies, where allocation of supervisory responsibility between supervisors from different disciplinary fields may lead to differing (and even competing) styles of support and guidance for the doctoral students.

Ethical awareness

The Standard requires that doctoral graduates demonstrate ethical awareness, and indeed, the requirement for understanding of professional conduct and research integrity is a fundamental aspect of any research programme. There are well-established national and international frameworks of definitions and regulations to guide ethical research conduct.

Disturbingly, the Report reveals that while most institutions have established processes for ethics approvals, few provide training programmes which cover more than the process of obtaining ethics approval. Even in these processes, there were reports of lack of clarity and consistency, and many bottlenecks reported. Of additional concern were reports that across some institutions, in fields of human and animal research, students' understanding of the regulations, and compliance, were inadequate.

How are we educating our doctoral students to grapple with issues of ethics and research integrity? Very few institutions or individuals reported focusing on in-depth consideration of ethical issues, as for example the philosophy of ethical research, fundamental ethical principles, and associated responsibilities related to the research itself, including benefit to communities, in research involving human subjects. A broad interpretation of ethical accountability would seek to reconcile private good with public good and the benefit to humankind, whether local, national or global.

Assessment of graduate attributes

Further to describing the graduate attributes, the Standard requires a description of systems for monitoring and assessing the progression towards their attainment, which is key to judging the attainment of knowledge and skills attributes by doctoral graduates. Few Self-Evaluation Reviews provided evidence of assessment beyond the examination of the thesis. We need to debate whether this is enough.

From the Report, the majority of universities clearly consider that successful examination of the thesis demonstrates attainment of at least some of the graduate attributes, as, for example, broad and expert knowledge, original contribution to the field, research methodology, reflection, rigorous academic writing, and critical and analytical thinking. The assumption is that doctoral graduates whose theses are accepted, and who have achieved a body of research which constitutes a contribution to knowledge, have developed the expected intellectual thinking skills – an assumption which may not be justified without clear evidence of intellectual depth.

Related to this, the successful publication of peer-reviewed articles emanating from the doctoral research and other research outputs such as patents, is regarded as evidence of original and innovative thinking contributing to the body of knowledge, and of development of an effective researcher with disciplinary and professional impact. The majority of universities expect the doctoral graduate to have published their results in peer-reviewed journals, and this undoubtedly demonstrates that the research has yielded useful research data, but not necessarily intellectual depth.

Many doctoral students are (or perceive that they are) expected, by supervisors and research leaders, to prioritise the generation of data and publication of the results, over the development of in-depth thinking about their subject area. The national and institutional systems of performance management, and incentivising of publication, drive this behaviour in a way that can lead to doctoral graduates who are highly skilled technically, but are not equipped with critical thinking ability.

It is clear from the information in the Report that not all universities have sufficiently developed strategies and mechanisms to purposefully build the graduate attributes into their doctoral qualifications, and to assess them. Critical and analytical thinking and problem-solving skills are not easily measured, or assessed, directly. Many universities indicated difficulty in identifying ways to verify doctoral graduates' ability to conduct research-related critical and analytical thinking.

If we wish to assure ourselves that our graduates are critical thinkers, our universities need to develop this capacity. We perhaps also need to consider methods of assessment, and a move beyond the conventional thesis examination and oral presentation to a requirement for active engagement with assessors and peers.

Recognised challenges

A number of challenges exist for universities in the provision of doctoral training, in addition to those described above. The South African university sector is highly diverse, and the Report recognises the importance of context, for different institutions. Historical changes, including mergers, restructuring, and realignment of qualifications, have led to an uneven doctoral education terrain. There are wide differences in numbers of students, academic workloads, expectations of research productivity, supervisory capacity, and institutional culture. Within this context, many institutions have been unable to create effective systems for high-level doctoral education, and doctoral students' experiences are highly variable across the sector.

Related to this, while some universities have well-established infrastructure, a few do not have adequate infrastructure to support doctoral education, and the cost of establishing adequate infrastructure is prohibitive. Some of these institutions are reliant on collaboration and partnerships to support the research needs of their doctoral students. This may not be altogether disadvantageous, as it can provide students with broader learning opportunities, and access to a more research-rich environment. However, it does not solve the problem of institutional capacity where the research-rich environment needs to be established in the longer term.

An overarching challenge, recognised widely, is the availability of funding for doctoral student bursaries. With insufficient funding, and a challenging socio-economic situation, many doctoral students are working part-time, which, while providing experience and exposure, may limit their capacity to engage adequately with the intellectual development expected of them. Related, limited availability of funds for mobility constrains the opportunities for broadening thinking through experiences in other countries or institutions.

Interventions to support intellectual development

Recognising the need for academic support for doctoral students' development, many institutions offer training and capacity-building programmes but, again, this is highly variable across the sector. Many universities offer research methodology training, writing skills development, etc., and host workshops, seminars and colloquia, and there are examples of excellent support programmes.

Writing Centres, dedicated to assisting postgraduate students, were mentioned in several reports, established to specifically develop competence in academic writing and communication, including relevant digital literacy skills appropriate for doctoral research. Writing Centres typically have consultants with language skills, but it is important that the staff need to have postgraduate education experience themselves, and they need to work in collaboration with supervisors, while not replacing the supervisors' role.

Some universities have Postgraduate Centres (or equivalent) where staff are dedicated to supporting postgraduate progress, with academic, disciplinary, and intellectual input. Others reported establishing Communities of Practice among doctoral students, and these can provide for peer support and intellectual discussion.

However, in most cases, these academic support activities are voluntary and participation in the programmes is inconsistent. Systems are required to monitor and evaluate the impact of academic support interventions, and these are not in place in all institutions. In particular, there is a need to identify specific activities that students should undertake, to assess their progress towards the graduate attributes, and to demonstrate attainment of the attributes on graduation. Such systems could be developed nationally, for the benefit of all institutions.

Mentoring programmes for doctoral students were not widely reported. Mentorship, where experienced individuals (who are academically qualified but not supervisors) offer the student advice, informal support, and wisdom, could provide additional guidance for doctoral students to develop intellectual depth.

A few universities reported exploring new approaches to supervision, including cohort models, supervisory panels, and interdisciplinary supervision teams. This is generally viewed as a valuable approach, especially where the doctoral studies are in inter-/multi-/transdisciplinary knowledge areas. Few reported developing approaches such as doctoral training centres[4], where structured doctoral education is being conducted collaboratively between institutions, and participants have opportunities to engage with, and learn from, a range of different stakeholders.

New approaches

Given the considerations described above, it is time to ask ourselves how we could improve the doctoral education we offer, to assure the academic and intellectual development of our doctoral students. One approach is to revise our programmes, and focus more deliberately on critical thinking.

In 2018, an article in *Nature* outlined a (then) new programme at Johns Hopkins University, intended to reform the training of doctoral students, to 'put the philosophy back into the Doctorate of Philosophy'[5]. The programme was designed to develop students' critical thinking abilities, and to improve their capacity to recognise rigour and to understand scientific integrity and

social responsibility aspects of research. The key point was that doctoral programmes tend to train students to be technically skilled, but not broad critical thinkers. Furthermore, students focusing on productivity would not, generally, consider the social impacts of their research or societal needs for research that leads to a better world. While it is acknowledged that the article[5] was written in the context of medical training, it may be argued that the same is true of doctoral programmes in many different disciplines. These concerns closely match those outlined in the present article, regarding the South African context.

The proposed solution[5] was based on the implementation of curriculated doctoral programmes with coursework modules incorporated, and the debate over the inclusion of modules aimed at developing the critical thinking, integrity and social responsibility aspects of the doctoral education. We, in South Africa, need to ask ourselves how we could incorporate learning to develop critical thinking, integrity and social responsibility in our doctoral programmes.

Perhaps what is needed, initially, is a revision of the academic activities in which our doctoral students are engaged, and the introduction of high-level curriculated doctoral modules to enable development of those broad and critical thinking abilities that we seek.

In South Africa, the PhD qualification explicitly excludes credit-bearing coursework modules, with the exception of the professional doctorate, over which there is some current debate. Currently, credit-bearing coursework and work-integrated learning are allowed in the professional doctorate, and the debate is around the challenge of ensuring quality and academic rigour in such doctoral-level coursework. The addition of curriculated training modules in general doctoral degree programmes may be a complex issue under the current national qualification framework, with respect to practical issues (such as impact on time required for completion of the programme) and the need to change the Higher Education Qualifications Sub-Framework regulations for doctoral qualifications. Nevertheless, there is an urgent need for alternative approaches to developing academic environments and processes that will impart and facilitate the development of 'thinking' capacity in doctoral students in South Africa.

Conclusion

We must debate whether South African institutions are doing enough to make doctoral programmes intellectually, academically, and philosophically directed, and whether there are sufficiently intensive opportunities for our doctoral candidates to engage in academic conversation and debate, and scholarly activities that add intellectual depth, and develop high-level critical thinking.

The National Review has shown that such activities are not central to many of the doctoral programmes offered at our universities. The result is that we have little certainty that our doctoral graduates do emerge as the broad and critical thinkers that we expect.

The National Report concludes with a set of recommendations, including that our institutions should:

- deliberately pursue awareness and integration of the graduate attributes in every doctoral programme;
- identify and clearly state assessment criteria and assessment tasks that doctoral students should complete in order to determine if the graduate attributes have been attained;
- consider fostering attributes such as critical citizenry and consciousness of social responsibility, with the notion of 'engaged research' and the doctoral qualification being seen as a 'public good'; and
- ensure that doctoral studies reflect global/international and regional benchmarks.

Engagement with the National Report, and its recommendations, should, at least, stimulate debate and discussion on how best to ensure that our doctoral education programmes achieve their purpose.

References

1. Council on Higher Education (CHE). Doctoral degrees national report. Pretoria: CHE; 2022. Available from: https://www.che.ac.za/sites/default/files/inline-files/CHE%20Doctoral%20Degrees%20National%20Reporte_0.pdf
2. Council on Higher Education (CHE). Higher Education Qualifications Sub-Framework: Qualification standard for doctoral degrees. Pretoria: CHE; 2018. Available from: https://nr-online-1.che.ac.za/html_documents/3.Doctoral%20Degrees%20Qualification%20Standard_Version%201.0_%20%20HEQC%20Approved%208%20November%202018%20(002).pdf
3. Faller F, Burton S, Kaniki A, Leitch A, Ntshoe I. Achieving doctorateness: Is South African higher education succeeding with graduate attributes? [submitted manuscript]; 2022.
4. Explore the different models of centres of excellence in doctoral training. Vitae News. 2012. Available from: https://www.vitae.ac.uk/news/vitae-news-2010-2015/explore-the-different-models-of-centres-of-excellence-in-doctoral-training
5. Bosch G. Train PhD students to be thinkers not just specialists. Nature. 2018;554:277. https://doi.org/10.1038/d41586-018-01853-1

AUTHORS:
Albert E. Manyuchi[1]
Matthew Chersich[2]
Coleen Vogel[1]
Caradee Y. Wright[3,4]
Ruwadzano Matsika[1]
Barend Erasmus[5]

AFFILIATIONS:
[1]Global Change Institute, University of the Witwatersrand, Johannesburg, South Africa
[2]Wits Reproductive Health and HIV Institute, University of the Witwatersrand, Johannesburg, South Africa
[3]Environmental and Health Research Unit, South African Medical Research Council, Pretoria, South Africa
[4]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa
[5]Office of the Dean, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Albert Manyuchi

EMAIL:
albertedgar.manyuchi@gmail.com

DATES:
Received: 21 Aug. 2021
Revised: 30 Aug. 2022
Accepted: 21 Sep. 2022
Published: 30 Nov. 2022

HOW TO CITE:
Manyuchi AE, Chersich M, Vogel C, Wright CY, Matsika R, Erasmus B. Extreme heat events, high ambient temperatures and human morbidity and mortality in Africa: A systematic review. S Afr J Sci. 2022;118(11/12), Art. #12047. https://doi.org/10.17159/sajs.2022/12047

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITOR:
Pascal Bessong

KEYWORDS:
Africa, heat-health plans and policy, heat-health interventions, heatwaves, high ambient temperature

FUNDING:
University of the Witwatersrand

Extreme heat events, high ambient temperatures and human morbidity and mortality in Africa: A systematic review

Temperature extremes vary across Africa. A continent-wide examination of the impacts of heat on health in Africa, and a synthesis of Africa-informed evidence is, however, lacking. A systematic review of articles published in peer-reviewed journals between January 1992 and April 2019 was conducted. To be eligible, articles had to be Africa-specific, in English, and focused on how heatwaves and high ambient temperatures affect morbidity and mortality. A secondary systematic analysis on policies and interventions comprising 17 studies was also conducted, and the findings synthesised together with those of the 20 primary studies. Eleven studies showed that high ambient temperatures and heat waves are linked with increased mortality rates in Africa. These linkages are characterised by complex, linear and non-linear (J or U) relationships. Eight of the nine primary studies of morbidity outcome reported that an increase in temperature was accompanied by raised disease incidence. Children and the elderly were the population groups most vulnerable to extreme heat exposure. Location-specific interventions and policy suggestions include developing early warning systems, creating heat-health plans, changing housing conditions and implementing heat-health awareness campaigns. In summary, this review demonstrates that, while heat-health relationships in Africa are complex, extreme temperatures are associated with high mortality and morbidity, especially amongst vulnerable populations. As temperatures increase across Africa, there is an urgent need to develop heat-health plans and implement interventions. Future studies must document intervention effectiveness and quantify the costs of action and inaction on extreme heat-related mortality and morbidity.

Significance:

- Empirical evidence shows that the relationship between heat and human health is complex in the African context. This complexity has implications for the development of interventions and policies for heat-health on the continent.
- This review is important for African policymakers, practitioners and others who support Africa's adaptation to climate change. Through this review, a compendium of Africa-specific and relevant empirical information is aggregated and made readily available to various interested and affected parties.

Introduction

Heatwaves and high ambient temperatures affect human health.[1,2] Biological, environmental, medical, socio-behavioural and geographical factors influence how extreme heat exposure affects morbidity and mortality.[3] The sub-groups most vulnerable to extreme temperatures include infants, the elderly, persons living with disabilities, persons on chronic medications, pregnant persons and outdoor workers.[4-8] Extreme heat-related diseases and deaths disproportionately affect the poor.[9]

Studies set in Africa which provide continent-specific insights on the effects and association between extreme temperatures on health are essential for planning actions to increase resilience, as well as programmes and policies for heat-health adaptation and mitigation.[10] The majority of studies on the relationship between extreme heat exposure, and morbidity and mortality have been carried out in high-income countries.[11-13] There are thus major gaps in knowledge about the effects of heatwaves on health outcomes among different sub-groups in low- and middle-income countries.[11,14-17] The few existing studies that address this topic in Africa largely focus on other continents and include selected African countries usually as an add-on. For example, Campbell et al.[18] undertook a review of global evidence which included no studies on Africa but generalised the conclusion from other continents to Africa. Green et al.[13] found few studies on Africa relative to other continents and observed that there are more studies on cities such as Brisbane and Phoenix than there are on cities in Africa.

To date, review studies examining Africa-wide effects and association between extreme heat events and high ambient temperatures on morbidity and mortality are lacking. Yet, observed climate trends show considerable increases in near surface temperatures over most parts of Africa in the past 50 to 100 years, with minimum temperatures warming rapidly relative to maximum temperatures.[19] In addition, future climate projections show that temperatures in Africa will rise much faster than the global average.[9]

The main objectives of this article were thus to provide an Africa-wide systematic review of evidence on the effects and association between heatwaves and high ambient temperatures on morbidity and mortality, to summarise the population groups most vulnerable to extreme heat exposure, and to explore the suggested heat-health interventions and policies to protect people on the African continent.

Methods

Scope of review

We conducted a systematic review of published peer-reviewed literature on the effects of extreme heat events and high ambient temperatures on morbidity and mortality in Africa. An adapted, realist review method was used. A realist review or synthesis broadly falls under systematic reviews. It takes a systematic approach to reviewing and synthesising available evidence.[20] A realist review attempts to provide answers to complex issues[21] by seeking detailed explanations of the issues[22]. It aims to provide practice and policy relevant evidence through a determination of what works in what circumstances, and for which particular persons, and why or how it works. It includes qualitative critical analyses which provide for depth rather than breadth.[23]

While the general principles of systematic reviews include 'the need for clear question(s), the need for transparency of methods and the use of wide-ranging, comprehensive searches to reduce the effects of publication bias'[24], a realist review encompasses tighter inclusion and exclusion criteria, and a smaller number of articles relative to other approaches[25]. A realist review is applicable for 'complex interventions where evidence of effect may be lacking'[26].

Rather than applying all the core features of the realist review method in their entirety, here we used an adapted realist synthesis. We retained the core features of a realist review, including developing a set of guiding questions and strict inclusion and exclusion criteria as well as conducting a wide-ranging and comprehensive search to minimise bias. However, instead of responding to the question of what works, our adapted realist review mainly responds to questions of effect and association between heat and heatwaves and human health.

Our adaptation of the realist review method is in agreement with Petticrew[24] who argues that systematic reviews should evolve from answering simple questions to respond to complex phenomena. We further undertook a secondary analysis of published data with less strict inclusion and exclusion criteria to examine policy and practice suggestions – something uncommon in realist synthesis. We considered the adapted realist review method most suitable to explore the complex subject of extreme heat-health relationships in data-scarce African settings.

In line with the adapted realist method, the following four background questions guided the review process: (1) What is the association between heatwaves and high ambient temperatures and morbidity and mortality in Africa? (2) What are the effects of extreme heat events and high ambient temperatures on specific population groups in Africa? (3) What are the suggested interventions to reduce morbidity and mortality from heatwaves and high ambient temperatures in Africa? Finally, (4) what are the policy propositions suggested to deal with heatwaves and high ambient temperature related morbidity and mortality? This study was approved by the Human Research Ethics Committee (Non-Medical) of the University of the Witwatersrand (protocol number H18/11/18). Although this review addresses heat exposure and human morbidity and mortality, we are not diminishing the major importance of heat impacts on morbidity and mortality in other animals and plants, which warrant equal attention.

Search strategy

Africa-focused, peer-reviewed studies published between 1 January 1992 and 30 April 2019 were reviewed. Reviewed studies focused on the effects and association between heatwaves and high ambient temperatures on morbidity and mortality in Africa. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed.[27] A topic search using the 'advanced search' option was performed within the Web of Knowledge, Scopus and PubMed electronic databases and applied the following phrases: Extreme heat events and mortality in Africa; Extreme heat events and morbidity in Africa; Heatwaves and mortality in Africa; Heatwaves and morbidity in Africa; High ambient temperatures and mortality in Africa; and High ambient temperatures and morbidity in Africa.

Eligibility criteria

All article titles and abstracts arising from the literature search were manually screened based on the inclusion and exclusion criteria presented in Table 1. Where the article titles and abstracts proved inadequate to determine relevance or eligibility, the full-text article was retrieved and assessed.

Table 1: Inclusion and exclusion criteria for systematic literature review

Inclusion criteria	Exclusion criteria
Articles published in English only	Articles published in languages other than English
Articles published from 1 January 1992 to 30 April 2019	Articles published prior to 1992 and after 1 May 2019
Peer-reviewed published articles with full-text access	Meetings, comments, abstracts, and inaccessible full-text articles
Articles whose focus was an African country, African countries or an African location	Global studies (not focused on Africa)
Articles whose main exposure of interest was extreme heat events (heatwaves) and high ambient temperatures	Articles whose main exposure of interest was not heatwaves or high ambient temperatures
Articles in which non-vector morbidity and mortality in human populations was analysed	Articles on vector-caused human morbidity and mortality

We also reviewed the reference lists of the retrieved full-text articles. From these lists, relevant references aligned to the inclusion criteria not initially identified were extracted and added. We included both qualitative and quantitative studies.

Data extraction

The first author extracted data using a standardised data collection form. The variables for data extraction included inter alia: the country or location of study; year of study; study design; study setting; study aims; study population; study outcome measures; geographical and temperature units (lags, temperature units); and results and conclusion of the study. Emphasis was placed on estimates of association (i.e. relative risk, attributable risk or percentage increase in morbidity and mortality) and effects on study population, especially sub-groups (by gender, age, cause of death, among others) obtained from tables, text descriptions and any other supplementary materials. The co-authors moderated the data extraction, resolved any differences, and provided any additional publications not located by the first author.

Synthesis of results and risk of bias

We collated extracted data and reviewed it using thematic analysis. In order to stress important points, we provide a synthesis of key results of some individual studies.

Additional secondary analyses of published data

We undertook an additional, secondary analysis encompassing global studies and Africa-specific studies that did not meet the review criteria, but where authors had suggested potential interventions and policies. For quality reasons, grey literature was excluded, and only peer-reviewed published literature was included in this additional policy-focused secondary analysis. For these studies, we outline the reported variables and summarise the interventions and policies suggested. While the studies did not meet all the study inclusion criteria, they nevertheless provided key insights (Supplementary table 1). The inclusion of the additional literature broadened the literature base from which we drew our findings. Supplementary table 1 shows the 17 reviewed articles included in the secondary analysis.

Results

Study selection and study characteristics

A flow diagram (Figure 1) shows the study selection informed by the inclusion and exclusion criteria.

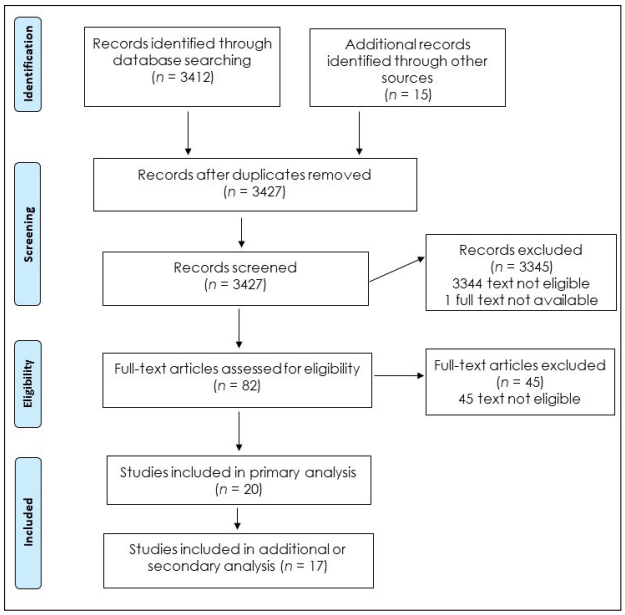


Figure 1: Flow diagram of the study selection strategy.

The study characteristics and findings for the 20 primary studies that were included in the review are summarised in Supplementary table 2. Based on the United Nations geoscheme classification of African sub-regions, the largest number of primary studies were set in the Southern Africa region (n=7), made up of six studies in South Africa and one in Botswana. There were six studies in Eastern Africa, with two in Kenya, two in Tanzania and one each in Mozambique and Zambia. Five studies were set in Western Africa, with two in Burkina Faso, two in Ghana and one in Senegal. Only one study in Cameroon was done in the Central (Middle) Africa region. Most studies were published between 2016 and 2019 (inclusive). The years in which data were collected in the primary studies ranged from 1974 to 2015.

Most studies (n=15) applied time-series methodology. The primary studies covered rural and urban areas of Africa and were either location specific, multi-location or nationwide. Among the primary studies, 11 adjusted for both seasonality and lag effect, 3 adjusted for seasonality but not lag effect, 4 adjusted for lag effect and not seasonality, and 2 did not report whether they adjusted for seasonality or lag effect. Several studies estimated the relative risk of association between daily mean temperature and daily mortality using lag strata.

Temperature data, including data sources

Diverse temperature measurements were used in the studies, including daily minimum and maximum temperatures, daily mean temperature, monthly mean temperature, and apparent air temperature (also called T_{app}). Most studies (n=17) focused on high ambient temperature, with only three assessing extreme heat events (heatwaves). Most temperature data were gathered from in-country National Meteorological and Hydrological Services, including those affiliated to the World Meteorological Organization. The National Meteorological and Hydrological Services providing temperature data were located in the study area or in close proximity. Some studies supplemented or combined in-country temperature data with those from external sources such as the National Oceanic and Atmospheric Administration of the United States of America, South Africa's Agriculture Research Council, the Climate Research Unit of the University of East Anglia, United Kingdom, and the Royal Netherlands Meteorological Institute's Climate Explorer website.

Health outcome measures

Of the 20 primary studies, 10 focused on mortality, 9 on morbidity and 1 covered both outcomes. Studies of mortality outcomes reported on all-cause mortality, as well as disease-specific mortality such as cardiovascular mortality. Most studies with mortality as an outcome measure ascertained cause of death through verbal autopsy questionnaires. Studies of morbidity outcome examined cholera and Ebola virus outbreaks (n=3), diarrhoea (n=4), acute respiratory infections (n=1) and cardiovascular diseases (n=1), among others.

Population groups examined

Most studies examined heat effects on specific age groups. The age categories were typically 0–4 years; 5–19 years; 20–49 years; 50–59 years and, 60 years and above. One study did not follow this age categorisation and its categories were from 40–64 years and then 65 years and above.[28] Four studies included only children or adolescents aged ≤18 years.[29-32]

Synthesis of key findings

(1) Extreme heat exposure and morbidity and mortality in Africa

Taking all the evidence together, it is clear that the association between extreme heat exposure and mortality in Africa is complex and varies considerably across settings (Supplementary table 2). Two studies found a linear association between high ambient temperatures and mortality.[33,34] Three studies found a non-linear relationship.[11,35,36] Three studies found a U-shaped relationship between heat exposure and mortality.[11,33,37] One reported a J-shaped temperature–mortality association for all-ages mortality and a U-shape for under-five mortality.[36] Most studies reported associations between temperature and all-cause mortality, but four described heat impacts on mortality related to non-communicable diseases such as cardiovascular disease.

Eight of the nine studies on temperature–morbidity relationships showed positive associations, that is, an increase in temperature resulted in an increased incidence of the conditions assessed. For example, a study reported that incidences of diarrhoea, respiratory infection, asthma, meningitis and malaria increased with an increase in temperature.[30] Six studies reported that rates of diarrhoeal diseases increased at higher temperatures, including two studies on cholera. Studies of morbidity outcome commonly pointed out that increased minimum temperature affects the prevalence of diseases in Africa. One study on heat and Ebola virus outbreaks found an association with cold, but not hot temperatures.

(2) Effects on population sub-groups

Most studies (n=11) whose outcome measure was mortality analysed the effects of extreme heat events and high ambient temperature on given population groups by age and gender (Supplementary table 2). The studies found that infants, children and young adolescents as well as the elderly (0–4 years; 5–12 years; 13–19 years and 50 years and above, respectively) were most vulnerable to deaths associated with exposure to extreme heat. In addition, higher mortality was found amongst male individuals than female individuals in one study assessing heat impacts on mortality from non-communicable diseases in Ghana.

Few studies of morbidity outcome analysed the effects of extreme heat events or high ambient temperature on population sub-groups. In one example, a study found increased incidences of diarrhoea, respiratory infections and malaria, particularly among male children compared to female children.[30] Two studies observed increased ambient temperatures were linked with increased diarrhoeal incidence among children under five years.[31,32] One study noted increased acute respiratory infections among children aged 18 years and younger.[29]

(3) Suggested interventions

Although we combined primary and secondary analysis results to elaborate the heat-health interventions focused on African localities, 21 out of the 37 studies explicitly offered Africa-appropriate suggestions, as shown in Table 2.

Table 2: Suggested interventions from reviewed literature

Author/s (year)	Summary of suggested intervention
Amegah et al.[42] (2016)	It is necessary to invest in meteorological services and strengthen health information systems to guarantee timely, up-to-date, and reliable data. It is also important to characterise the groups vulnerable to heat and develop plans to strengthen resilience in these groups. Profiling of socio-economic differentials in the temperature-health relationships must inform preventative actions.
Dukić et al.[38] (2012)	Interventions aimed at reducing temperature, dust, smoke and carbon dioxide exposure or increasing relative humidity and ventilation via some household intervention mechanism may reduce the public health burden of meningitis outbreaks, especially during large outbreaks.
Trærup et al.[39] (2011)	There is need for health education to raise awareness of cholera in Tanzania. Furthermore, disease surveillance, early diagnosis and treatment are essential. It is necessary to implement programmes that improve access to clear water and sanitation.
Oloukoi et al.[44] (2014)	It is important to take urgent action to provide corrective and preventative adaptation measures to combat climate-health risks in Oke-Ogun region. Primary prevention interventions include providing access to safe drinking water, while secondary prevention entails strengthening disease surveillance programmes and early response to disease outbreaks. Supply of medical personnel and systematising improved hygiene behaviour at household level is important as both a preventative and corrective intervention.
McMichael et al.[35] (2008)	It is essential to minimise unplanned rapid urban development because this has adverse effects on sanitation, air pollution and housing, which affect population vulnerability to extremes of heat exposure. It may be important to develop heat-health warning systems.
Grace et al.[53] (2012)	Considering that there is an association between malnutrition and water source, malnutrition and floor type and malnutrition and education in Kenya, it is important to invest in improved water delivery infrastructure to reduce malnutrition. Because climate change can increase child stunting rates, investments must be made to ensure all households have access to culturally relevant and nutritional food.
Bandyopadhyay et al.[54] (2012)	To reduce diarrhoea, it is important to reduce water shortages, promote handwashing, and improve access to toilets and better health services, especially among poor households in Africa. Early warning systems that facilitate a preventative approach are important. It is also important to develop spatial disease maps that show the prevalence of diarrhoea by region and to identify areas where the risk of diarrhoea from water scarcity and elevated temperature is high.
Green et al.[13] (2019)	It is imperative to improve meteorological and health data, especially heat-health data systems by deploying low-cost sensors in low-income countries. Conditional cash transfers can be an intervention to reduce population vulnerability to heat exposure. Heatwave early warning systems must be developed. Research on actions that reduce mortality during heatwaves is essential.
MacVicar et al.[15] (2017)	In rural Uganda, it may be important to provide individuals in the third trimester of pregnancy with additional nutritional resources, especially during the dry season. It is also important to monitor changes in seasonal temperature and provide targeted nutritional interventions for those who are pregnant during the 'hot' season.
Azongo et al.[11] (2012)	Improve ventilation and sleeping arrangements during hot weather conditions.
Thompson et al.[30] (2012)	Climate-health responses for children should happen at the family, community, local municipality, NGO and international agency levels. Awareness campaigns on health care of children must be emphasised. Children's healthcare delivery must be improved. As poor health outcomes among children are related to poverty, job creation is important. Good environmental practices must be practices in communities including regular environmental sanitation. Environmental education programmes must be implemented, including those focusing on preventing wild veld fires. Tree planting campaigns must also be carried out in communities. Stakeholders, especially women and children, must be involved in child health programmes and awareness of climate change impacts raised in communities.
Alexander et al.[55] (2013)	Locally applied interventions that are sustainable must include socio-cultural considerations into public health planning.
Heunis et al.[43] (1995)	It is important to issue early warning weather forecasts to susceptible sectors of the population.
Tchidjou et al.[29] (2010)	The high frequency of hospitalisation from acute respiratory infections suggests that influenza vaccination campaigns should be implemented, considering the seasonality in Cameroon.
Reyburn et al.[41] (2011)	It is important to create an early warning system for cholera in Zanzibar because this will give public health authorities time to mobilise staff and equipment to respond to an outbreak. An early warning system can facilitate mass oral cholera vaccination.
Fernández et al.[56] (2009)	A cholera early warning system is imperative for sub-Saharan Africa.
Ng et al.[45] (2014)	Support in case detection and reporting on Ebola virus outbreak countries must be supported. Also, longitudinal serological and virological surveillance studies must be implemented to better understand Ebola virus transmission patterns.
Bunker et al.[57] (2017)	It is important to develop early preventive measures to curb heat-associated non-communicable disease deaths in Nouna, Burkina Faso.
Horn et al.[40] (2018)	Increasing access to oral rehydration in local health centres and increasing education on appropriate use and handling of water and on sanitation practices can help reduce transmission of diarrhoeal pathogens and consequently the outbreaks of diarrhoea diseases in some areas in Mozambique. Furthermore, an early warning system for conditions that favour diarrhoea disease should be developed.
Thiam et al.[32] (2017)	There is a need for effective preventive measures to reduce the high burden of diarrhoea in the health district of Mbour, Senegal. Health intervention programmes in the cold dry season and in the rainy season focusing on morbidity control and prevention should be launched, particularly in urban settings where diarrhoea is most common, to reduce the incidence of diarrhoea in this context of climatic variability.
Scovronick and Armstrong[14] (2012)	Interventions must focus on redesigning human settlements in informal settlements.
Egondi et al.[36] (2012)	The fact that children and the elderly are susceptible to weather mortality signifies the need for weather-mortality efforts such as proper housing and clothing.
Egondi et al.[58] (2015)	It is important to create better targeted awareness campaigns to reduce the health burden from temperature exposure.

Table 2 shows that a few studies suggested either preventative or treatment-related interventions at individual, household and community levels.[30,32,38] Although most studies were location specific, most of the suggestions made were likely to be applicable to other similar localities. Several individual- and household-level preventative interventions were identified, such as increased educational campaigns aimed at raising awareness about the effects and association between extreme heat events and human health.[29,30,39] Another example is seasonal forecast-based vaccination campaigns aimed at reducing hospital admissions from acute respiratory infection and reducing exposure to extreme temperature particularly among vulnerable groups such as young children and the elderly.[11,29]

Importantly, because there is a high mortality associated with meningitis during the dry season in the meningitis belt in Africa, the main suggestions were to redesign housing structures so that they have proper ventilation, to change sleeping arrangements as well as to reduce dust, smoke and carbon dioxide exposure.[38] It is worth noting that better ventilation may not necessarily mean providing air conditioning in the African context but constructing structures with larger and well-positioned windows and doors, for example.[14] In order to reduce smoke at household level, cooking stoves with better combustion properties were suggested.[38]

A community-level preventative measure suggested was the proper allocation of resources, especially to guarantee that vulnerable persons are protected against the effects of extreme heat. This suggestion was focused on governments as well as other persons with authority to distribute resources in Africa. In fact, reviewed studies focusing on South Africa – where there are extensive informal settlements – generally pointed out that informal settlements must be replaced, and new heat-adapted housing structures should be developed.[14] The same study pointed out that traditional, round thatch huts – especially in the rural areas of Africa – should also be modified for adaptation to extreme heat events and high ambient temperatures.[14]

Studies focused on diarrhoea and cholera incidences suggested the importance of strengthening household and community interventions for preventing weather-induced diarrhoea and cholera diseases in Africa.[32,39-41] One study pointed out the potential benefits of conditional cash transfers to deal with vulnerability related to heat exposure[12], while another alluded to the use of clothing suitable to hot climates[36].

Early warning system interventions were suggested several times, to ensure that the African populace is informed of impending extreme heat events before they occur.[13,35,40-43] While being informed may not directly translate to action, early warning systems might enable the populations that will be affected by exposure to extreme heat to prepare in advance. This suggestion implicitly calls for the strengthening of climate information services for health on the continent. The few reviewed studies offering policy suggestions posited that better climate data are essential for better health-focused responses. Indeed, quality data from all relevant stakeholders and actors are essential for heat-health planning efforts required. Surveillance systems were suggested in several studies reviewed.[39,44,45]

(4) Suggested policies

Few studies proffered concrete policy suggestions useful for extreme heat-health policymaking in Africa. Only nine out of 37 studies gave some explicit policy suggestions, as shown in Table 3.

Three of these studies implicitly pointed out the importance of developing heat-health plans in Africa.[14,35,44] However, not much detail was provided on the content and implementation of such plans. The cautionary advice provided was, as Africa comprises widely diverse localities, developing a 'one-size-fits-all' heat plan could be ineffective. The findings of Scott et al.[46] on urban heat islands in Nairobi were informative as they emphasised location-specific factors when assessing morbidity and mortality in Africa, particularly demonstrating that studies with large spatial coverage might not accurately illuminate the range of heat-morbidity and -mortality relationships. Rather, more localised heat plans should be an imperative.

Another key suggestion is for authorities to formulate policies that target vulnerable populations that would be affected by exposure to extreme temperatures.[30] Heat plans must explicitly outline the different levels and categories of vulnerability in the communities they cover.[35] Therefore, data that explicitly delineate vulnerable populations are essential for crafting an effective heat plan.

Table 3: Key policy suggestions from reviewed literature

Author/s (year)	Summary of key policy suggestion
Amegah et al.[42] (2016)	Sub-Saharan African governments must take urgent steps to remove the bottlenecks to accessing institutional data for research.
Trærup et al.[39] (2011)	Integrated socio-economic, climate and health policies that facilitate provision of clean water and sanitation may reduce the costs of responding to outbreaks of cholera in Tanzania.
Oloukoi et al.[44] (2014)	Considering that local populations of Oke-Ogun region of Nigeria use indigenous therapy encompassing use of herbs and ash for residential vaccination against measles in particular, the incorporation of indigenous knowledge into climate change policies may lead to creation of effective adaptation strategies that are cost effective, participatory and sustainable. Policies that promote integration of health outcomes into development policies, especially focusing on prevention of ill-health are imperative. Policies that promote involvement of various stakeholders and facilitate information sharing on the health impacts of climate change are necessary for Nigeria and Africa.
McMichael et al.[35] (2008)	Effective preventive or adaptive policies based on an understanding of demographic, social and ecological determinants, and sensitivities of the response pattern of the population of concern are important.
Bandyopadhyay et al.[54] (2012)	Policies that improve nutritional status, sanitation, and parental education and provide better access to health services are important to safeguard against diarrhoea.
Green et al.[13] (2019)	It is imperative to create policies that promote health networks for heat-health research and invest financial and human resources in these.
Scovronick and Armstrong[14] (2012)	To deal with heat, it is imperative to develop a development policy that prioritises replacement of informal housing compared to one that emphasises replacement of traditional dwellings in South Africa.
Thompson et al.[30] (2012)	It is important to create climate-health policies with a focus on children. The municipalities in Limpopo must provide the necessary legislative and administrative framework for good environmental practices such as regular environmental sanitation. It is important to create policies to reduce deforestation and soil erosion, especially in high-density areas of uneven terrain. Policies for mapping disaster risk areas and plans to target these areas must be developed.
Alexander et al.[55] (2013)	Locally acceptable and culturally appropriate public health strategies for reducing population vulnerability to climate change related diarrhoea disease must be developed.

Discussion

This review summarises the evidence published between January 1992 and April 2019 on the effects of extreme heat events and high ambient temperatures on mortality and morbidity in Africa. To our knowledge, this is the most up-to-date Africa-focused systematic review on this subject, notwithstanding the existence of other studies that have included African countries in reviews that cover studies across the globe.[13,47] A previous review focused on countries in sub-Saharan Africa and excluded studies in North Africa[42], and a more recent study included only 10 studies from Africa among many studies from other continents[13].

The general pattern among studies that explored the heat-health relationship in Africa was exclusionary (limited in coverage) and

non-exhaustive (referencing the continent in passing). Our review encompassed 20 primary studies and 17 secondary studies, and importantly highlighted Africa-specific nuances on mortality and morbidity arising from extreme heat exposure. This review underscores the fact that Africa has particular climatic and population health features that demand a focused and thorough exploration.

Overall, the review confirmed the effects of extreme heat on mortality and morbidity in Africa. This finding is in line with previous related reviews in Africa and globally. For instance, a study reported the dominance of heat effects in studies conducted in sub-Saharan Africa, with heat-related mortality observed in Burkina Faso, Ghana and Kenya.[47] Even though Green et al.[13] included studies from other continents, they similarly point out that

> *the vast majority of studies found a positive association between heat and morbidity and mortality, and only nine either found no association or a negative association between the heat exposure and the health outcome of interest.* (p. 84)

Not many reviews provide unequivocal, continent-wide evidence on the effects of extreme heat exposure on mortality and morbidity.

This review further points out that the association between extreme heat and mortality or morbidity is complex. Indeed, the heat-mortality and heat-morbidity relationships were found to be linear or non-linear, with the latter showing U- and J-shaped relationships. We suggest that the complexity in these relationships may result from the uneven geographic spatial distribution of climate elements and the epidemiological heterogeneity of the continent. The complex relationship found demands further research to explore other factors, including inter alia whether the methodology employed in the studies, the sub-populations investigated, and other independent variables directly or indirectly influence the relationship.

The evidence indicates that although extreme heat effects affect persons of all age groups, the African population groups most affected by exposure to extreme heat were infants (0–4 years), children (5–19 years) and the elderly (over 50 years). Similar findings have been reported in studies from South Asia where heat-mortality effects are more pronounced among children[48-50], and in Latin America and China where heat-mortality effects were observed among the elderly[51,52].

A few studies suggested some practical interventions and policies to deal with the effects of extreme heat events and high ambient temperatures in Africa. Many of these suggestions pertained to preventative and treatment interventions, as well as polices for early warning systems and heat-health plans. The suggested interventions and policies, however, lack depth and detail. More intervention and policy-oriented studies that focus on these study areas are needed. For studies to offer useful programme and policy suggestions, more Africa-specific intervention projects must be carried out.

There are some limitations to this review. The main limitation is that we did not evaluate the quality or potential biases of studies included in this review. Heterogeneity of included studies made it difficult to consider doing a meta-analysis for this review. The allocation of studies into those fitting primary or secondary analysis was not error-free. The general limitation of the adapted realist review method, which is that it may not be reproducible and transparent, applies to this review.[26] Some key concepts with fluid definitions such as extreme heat and heatwaves were not explicitly defined, especially within the context of the continent which has predominantly warm/hot tropical and subtropical climates. In addition, the causal effects of heat impacts on health, for example, physiological or cardio-metabolic stress versus communicable diseases, were not fully explored. However, some of these concepts are complex to explain robustly without oversimplifying them and making erroneous attributions. These are areas for future studies to explore.

Conclusion

While there is a growing body of evidence suggesting that Africa will be disproportionately affected by climate change, the effects of extreme temperature changes on health on the continent have not been explored in detail to date. This systematic review established the effects and association between extreme heat and high ambient temperatures on mortality and morbidity in Africa and explored the population sub-groups most vulnerable. The evidence shows that the extreme heat-health relationship is complex and confirmed the effects of extreme temperature on mortality and morbidity, especially among children and the elderly.

African countries have a substantial mortality and disease burden. It is imperative to reduce extreme temperature related deaths and diseases. Future climate projections show increased temperatures across many countries on the continent. Therefore, policy and programmatic measures that curb future heat-health mortality and morbidities are urgently needed. Research that explores the economic costs of proactively and reactively acting on heat-health morbidities and mortalities must be carried out to inform policymakers and practitioners.

Suggested interventions and policies to deal with the effects of exposure to extreme heat in Africa were also sought in the review. Unfortunately, only a few of the reviewed studies provided detailed heat-health policy and programme suggestions. Further intervention and policy-focused research is encouraged. Large-scale, Africa-specific transdisciplinary studies that examine extreme heat-health relationships at sub-regional or continental scales are recommended.

Acknowledgements

A.E.M. received a bursary from the Global Change Institute (GCI) of the University of the Witwatersrand for this research, which was carried out as part of his PhD research. The GCI funding partners played no role in the design, execution or writing of this paper.

Competing interests

We have no competing interests to declare.

Authors' contributions

A.E.M. designed the study and performed the data extraction, and wrote the first and final drafts, taking into consideration the different sections and reviews by the co-authors. M.C. reviewed the data collection instruments, guided data collection and reviewed the first draft and updated the methodology section. C.Y.W., C.V. and R.M. reviewed the drafts and C.Y.W. and R.M. worked on all tables. B.E. helped with the analysis and review of extracted articles and reviewed the final manuscript.

References

1. Huang C, Barnett AG, Wang X, Vaneckova P, Fitzgerald G, Tong S. Projecting future heat-related mortality under climate change scenarios: A systematic review. Environ Health Perspect. 2011;119(12):1681–1690. https://doi.org/10.1289/ehp.1103456
2. Zeng Q, Li G, Cui Y, Jiang G, Pan X. Estimating temperature-mortality exposure-response relationships and optimum ambient temperature at the multi-city level of China. Int J Environ Res Public Health. 2016;13(3):279–290. https://doi.org/10.3390/ijerph13030279
3. Xu Y, Dadvand P, Barrera-Gómez J, Sartini C, Marí-Dell'Olmo M, Borrell C, et al. Differences on the effect of heat waves on mortality by sociodemographic and urban landscape characteristics. J Epidemiol Community Health. 2013;67(6):519–525. https://doi.org/10.1136/jech-2012-201899
4. Oudin Åström D, Bertil F, Joacim R. Heat wave impact on morbidity and mortality in the elderly population: A review of recent studies. Maturitas. 2011;69(2):99–105. https://doi.org/10.1016/j.maturitas.2011.03.008
5. Benmarhnia T, Deguen S, Kaufman JS, Smargiassi A, Benmarhnia T, Deguen S, et al. Vulnerability to heat-related mortality: A systematic review, meta-analysis, and meta-regression analysis. Epidemiology. 2015;26(6):781–793. https://doi.org/10.1097/EDE.0000000000000375
6. Costello A, Abbas M, Allen A, Ball S, Patterson C. Managing the health effects of climate change. Lancet. 2009;373(9676):1693–1733. https://doi.org/10.1016/S0140-6736(09)60935-1
7. Loughnan M, Tapper NJ, Phan T, Lynch K, McInnes J. A spatial vulnerability analysis of urban populations during extreme heat events in Australian capital cities [document on the Internet]. c2013 [cited 2021 Aug 20]. Available from: https://nccarf.edu.au/wp-content/uploads/2019/03/Loughnan_2013_Spatial_vulnerability_analysis.pdf

8. Yin Q, Wang J. The association between consecutive days' heat wave and cardiovascular disease mortality in Beijing, China. BMC Public Health. 2017;17(1):1–9. https://doi.org/10.1186/s12889-017-4129-7

9. Hoegh-Guldberg O, Jacob D, Taylor M, Bindi M, Brown S, Camilloni I, et al. Impacts of 1.5ºC global warming on natural and human systems. In: Masson-Delmotte V, Zhai P, Pörtner H-O, Roberts D, Skea J, Shukla PR, et al., editors. Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Cambridge, UK: Cambridge University Press; 2018. p. 175–312. https://doi.org/10.1017/9781009157940.005

10. Fraser A, Leck H, Parnell S, Pelling M, Brown D, Lwasa S. Meeting the challenge of risk-sensitive and resilient urban development in sub-Saharan Africa: Directions for future research and practice. Int J Disaster Risk Reduct. 2017;2:106–109. https://doi.org/10.1016/j.ijdrr.2017.10.001

11. Azongo DK, Awine T, Wak G, Binka FN, Oduro AR. A time series analysis of weather variability and all-cause mortality in the Kasena-Nankana districts of Northern Ghana, 1995-2010. Glob Health Action. 2012;5:14–22. https://doi.org/10.3402/gha.v5i0.19073

12. Wichmann J. Heat effects of ambient apparent temperature on all-cause mortality in Cape Town, Durban and Johannesburg, South Africa: 2006–2010. Sci Total Environ. 2017;587–588:266–272. https://doi.org/10.1016/j.scitotenv.2017.02.135

13. Green H, Bailey J, Schwarz L, Vanos J, Ebi K, Benmarhnia T. Impact of heat on mortality and morbidity in low and middle income countries: A review of the epidemiological evidence and considerations for future research. Environ Res. 2019;171:80–91. https://doi.org/10.1016/j.envres.2019.01.010

14. Scovronick N, Armstrong B. The impact of housing type on temperature-related mortality in South Africa, 1996-2015. Environ Res. 2012;113:46–51. http://dx.doi.org/10.1016/j.envres.2012.01.004

15. MacVicar S, Berrang-Ford L, Harper S, Huang Y, Bambaiha DN, Yang S. Whether weather matters: Evidence of association between in utero meteorological exposures and foetal growth among Indigenous and non-Indigenous mothers in rural Uganda. PLoS ONE. 2017;12(6), e0179010. https://doi.org/10.1371/journal.pone.0179010

16. Medina-Ramón M, Zanobetti A, Cavanagh DP, Schwartz J. Extreme temperatures and mortality: Assessing effect modification by personal characteristics and specific cause of death in a multi-city case-only analysis. Environ Health Perspect. 2006;114(9):1331–1336. https://doi.org/10.1289/ehp.9074

17. Michelozzi P, Accetta G, De Sario M, D'Ippoliti D, Marino C, Baccini M, et al. High temperature and hospitalizations for cardiovascular and respiratory causes in 12 European cities. Am J Respir Crit Care Med. 2009;179(5):383–389. https://doi.org/10.1164/rccm.200802-217OC

18. Campbell S, Remenyi TA, White CJ, Johnston FH. Heatwave and health impact research: A global review. Health Place. 2018;53:210–218. https://doi.org/10.1016/j.healthplace.2018.08.017

19. Niang I, Ruppel OC, Abdrabo MA, Essel A, Lennard C, Padgham J, et al. Africa. In: Barros VR, Field CB, Dokken DJ, Mastrandrea MD, Mach KJ, Bilir TE, et al., editors. Climate change 2014: Impacts, adaptation, and vulnerability. Part B: Regional aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press; 2014. p. 1199–1265. https://doi.org/10.1017/CBO9781107415386.002

20. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: Meta-narrative reviews. BMC Med. 2013;11(20):1–15. https://doi.org/10.1186/1741-7015-11-20

21. Greenhalgh T, Wong G, Westhorp G, Pawson R, Wong G. RAMESES publication standards: Realist syntheses. BMC Med. 2013;11(21):1–14. https://doi.org/10.1186/1741-7015-11-21

22. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review – a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy. 2005;10(1): 21–34. https://doi.org/10.1258/1355819054308530

23. Thompson HE, Berrang-Ford L, Ford JD. Climate change and food security in sub-Saharan Africa: A systematic literature review. Sustainability. 2010;2(8):2719–2733. https://doi.org/10.3390/su2082719

24. Petticrew M. Time to rethink the systematic review catechism? Moving from 'what works' to 'what happens'. Syst Rev. 2015;4:36. https://doi.org/10.1186/s13643-015-0027-1

25. Rycroft-Malone J, McCormack B, Hutchinson AM, DeCorby K, Bucknall TK, Kent B, et al. Realist synthesis: Illustrating the method for implementation research. Implement Sci. 2012;7(1):1–10. https://doi.org/10.1186/1748-5908-7-33

26. Petticrew M, Rehfuess E, Noyes J, Higgins JPT, Mayhew A, Pantoja T, et al. Synthesizing evidence on complex interventions: How meta-analytical, qualitative, and mixed-method approaches can contribute. J Clin Epidemiol. 2013;66(11):1230–1243. http://dx.doi.org/10.1016/j.jclinepi.2013.06.005

27. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. BMJ. 2009;339(7716):332–336. http://dx.doi.org/doi:10.1136/bmj.b2535

28. Kynast-Wolf G, Preuß M, Sié A, Kouyaté B, Becher H. Seasonal patterns of cardiovascular disease mortality of adults in Burkina Faso, West Africa. Trop Med Int Health. 2010;15(9):1082–1089. http://dx.doi.org/10.1111/j.1365-3156.2010.02586.x

29. Tchidjou HK, Vescio F, Boros S, Guemkam G, Minka E, Lobe M, et al. Seasonal pattern of hospitalization from acute respiratory infections in Yaoundé, Cameroon. J Trop Pediatr. 2010;56(5):317–320. https://doi.org/10.1093/tropej/fmp127

30. Thompson AA, Matamale L, Kharidza SD. Impact of climate change on children's health in Limpopo province, South Africa. Int J Environ Res Public Health. 2012;9(3):831–854. https://doi.org/10.3390/ijerph9030831

31. Musengimana G, Mukinda FK, Machekano R, Mahomed H. Temperature variability and occurrence of diarrhoea in children under five-years-old in Cape Town metropolitan sub-districts. Int J Environ Res Public Health. 2016;13(9):1–12. https://doi.org/10.3390/ijerph13090859

32. Thiam S, Diène AN, Sy I, Winkler MS, Schindler C, Ndione JA, et al. Association between childhood diarrhoeal incidence and climatic factors in urban and rural settings in the health district of Mbour, Senegal. Int J Environ Res Public Health. 2017;14(9):1–16. https://doi.org/10.3390/ijerph14091049

33. Diboulo E, Sié A, Rocklöv J, Niamba L, Yé M, Bagagnan C, et al. Weather and mortality: A 10 year retrospective analysis of the Nouna health and demographic surveillance system, Burkina Faso. Glob Health Action. 2012;5:6–13. https://doi.org/10.3402/gha.v5i0.19078

34. Mrema S, Shamte A, Selemani M, Masanja H. The influence of weather on mortality in rural Tanzania: A time-series analysis 1999-2010. Glob Health Action. 2012;5:33–43. https://doi.org/10.3402/gha.v5i0.19068

35. McMichael AJ, Wilkinson P, Kovats RS, Pattenden S, Hajat S, Armstrong B, et al. International study of temperature, heat and urban mortality: The 'ISOTHURM' project. Int J Epidemiol. 2008;37(5):1121–1131. https://doi.org/10.1093/ije/dyn086

36. Egondi T, Kyobutungi C, Kovats S, Muindi K, Ettarh R, Rocklöv J. Time-series analysis of weather and mortality patterns in Nairobi's informal settlements. Glob Health Action. 2012;5:23–32. https://doi.org/10.3402/gha.v5i0.19065

37. Scovronick N, Sera F, Acquaotta F, Garzena D, Fratianni S, Wright CY, et al. The association between ambient temperature and mortality in South Africa: A time-series analysis. Environ Res. 2018;161:229–235. https://doi.org/10.1016/j.envres.2017.11.001

38. Dukić V, Hayden M, Forgor AA, Hopson T, Akweongo P, Hodgson A, et al. The role of weather in meningitis outbreaks in Navrongo, Ghana: A generalized additive modeling approach. J Agric Biol Environ Stat. 2012;17(3):442–460. https://doi.org/10.1007/s13253-012-0095-9

39. Trærup SLM, Ortiz RA, Markandya A. The costs of climate change: A study of cholera in Tanzania. Int J Environ Res Public Health. 2011;8(12):4386–4405. https://doi.org/10.3390/ijerph8124386

40. Horn LM, Hajat A, Sheppard L, Quinn C, Colborn J, Zermoglio MF, et al. Association between precipitation and diarrheal disease in Mozambique. Int J Environ Res Public Health. 2018;15(4), Art. #709. https://doi.org/10.3390/ijerph15040709

41. Reyburn R, Kim DR, Emch M, Khatib A, Von Seidlein L, Ali M. Climate variability and the outbreaks of cholera in Zanzibar, East Africa: A time series analysis. Am J Trop Med Hyg. 2011;84(6):862–869. https://doi.org/10.4269/ajtmh.2011.10-0277

42. Amegah AK, Rezza G, Jaakkola JJK. Temperature-related morbidity and mortality in sub-Saharan Africa: A systematic review of the empirical evidence. Environ Int. 2016;91:133–149. https://doi.org/10.1016/j.envint.2016.02.027

43. Heunis JC, Olivier J, Bourne DE. Short-term relationships between winter temperatures and cardiac disease mortality in Cape Town. S Afr Med J.1995;85(10):965–967.

44. Oloukoi G, Bob U, Jaggernath J. Perception and trends of associated health risks with seasonal climate variation in Oke-Ogun region, Nigeria. 2014;25:47–55. http://dx.doi.org/10.1016/j.healthplace.2013.09.009

45. Ng S, Basta NE, Cowling BJ. Association between temperature, humidity and ebolavirus disease outbreaks in Africa, 1976 to 2014. Euro Surveill. 2014;19(35), Art. #20892. http://dx.doi.org/10.2807/1560-7917.es2014.19.35.20892

46. Scott AA, Misiani H, Okoth J, Jordan A, Gohlke J, Ouma G, et al. Temperature and heat in informal settlements in Nairobi. PLoS ONE. 2017;12(11), e0187300. https://doi.org/10.1371/journal.pone.0187300

47. Burkart K, Khan MMH, Schneider A, Breitner S, Langner M, Krämer A, et al. The effects of season and meteorology on human mortality in tropical climates: A systematic review. Trans R Soc Trop Med Hyg. 2014;108(7):393–401. https://doi.org/10.1093/trstmh/tru055

48. Hashizume M, Armstrong B, Hajat S, Wagatsuma Y, Faruque ASG, Hayashi T, et al. Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: Effects and vulnerable groups. Int J Epidemiol. 2007;36(5):1030–1037.

49. Lindeboom W, Alam N, Begum D, Streatfield PK. The association of meteorological factors and mortality in rural Bangladesh, 1983-2009. Glob Health Action. 2012;5:61–73. https://doi.org/10.1093/ije/dym148

50. Ingole V, Kovats S, Schumann B, Hajat S, Rocklöv J, Juvekar S, et al. Socioenvironmental factors associated with heat and cold-related mortality in Vadu HDSS, western India: A population-based case-crossover study. Int J Biometeorol. 2017;61(10):1797–1804. https://doi.org/10.1007/s00484-017-1363-8

51. Wu W, Xiao Y, Li G, Zeng W, Lin H, Rutherford S, et al. Temperature-mortality relationship in four subtropical Chinese cities: A time-series study using a distributed lag non-linear model. Sci Total Environ. 2013;449:355–362. http://dx.doi.org/10.1016/j.scitotenv.2013.01.090

52. Goggins WB, Chan EY, Yang C, Chong M. Associations between mortality and meteorological and pollutant variables during the cool season in two Asian cities with sub-tropical climates: Hong Kong and Taipei. Environ Health. 2013;12(1), Art. #59. https://doi.org/10.1186/1476-069X-12-59

53. Grace K, Davenport F, Funk C, Lerner AM. Child malnutrition and climate in sub-Saharan Africa: An analysis of recent trends in Kenya. Appl Geogr. 2012;35(1–2):405–413. http://dx.doi.org/10.1016/j.apgeog.2012.06.017

54. Bandyopadhyay S, Kanji S, Wang L. The impact of rainfall and temperature variation on diarrheal prevalence in sub-Saharan Africa. Appl Geogr. 2012;33(1):63–72. http://dx.doi.org/10.1016/j.apgeog.2011.07.017

55. Alexander KA, Carzolio M, Goodin D, Vance E. Climate change is likely to worsen the public health threat of diarrheal disease in Botswana. Int J Environ Res Public Health. 2013;10(4):1202–1230. https://doi.org/10.3390/ijerph10041202

56. Fernández LMÁ, Bauernfeind A, Jiménez JD, Gil CL, El Omeiri N, Guibert DH. Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003-2006: Analysis of a time series. Trans R Soc Trop Med Hyg. 2009;103(2):137–143. https://doi.org/10.1016/j.trstmh.2008.07.017

57. Bunker A, Sewe MO, Sié A, Rocklöv J, Sauerborn R. Excess burden of non-communicable disease years of life lost from heat in rural Burkina Faso: A time series analysis of the years 2000-2010. BMJ Open. 2017;7(11), e018068. https://doi.org/10.1136/bmjopen-2017-018068

58. Egondi T, Kyobutungi C, Rocklöv J. Temperature variation and heat wave and cold spell impacts on years of life lost among the urban poor population of Nairobi, Kenya. Int J Environ Res Public Health. 2015;12:2735–2748. http://dx.doi.org/10.3390/ijerph120302735

AUTHOR:
Solomon R. Benatar[1,2]

AFFILIATIONS:
[1]Emeritus Professor of Medicine and Senior Scholar, University of Cape Town, Cape Town, South Africa
[2]Adjunct Professor, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

CORRESPONDENCE TO:
Solomon Benatar

EMAIL:
Solomon.benatar@uct.ac.za

DATES:
Received: 30 Jan. 2022
Revised: 02 Aug. 2022
Accepted: 06 Oct. 2022
Published: 30 Nov. 2022

HOW TO CITE:
Benatar SR. COVID-19, global health and climate change: Causes and convergences. S Afr J Sci. 2022;118(11/12), Art. #13165. https://doi.org/10.17159/sajs.2022/13165

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITOR:
Leslie Swartz

KEYWORDS:
global health, COVID-19, climate change, multiple instabilities

FUNDING:
University of Cape Town

COVID-19, global health and climate change: Causes and convergences

Despite massive global economic growth and advances in science and medicine with spectacular aggregate and individual improvements in health and life expectancy over the past century, the world has now become severely unstable in multiple domains – biological, sociological, political, ecological, economic, and health care. These pervasive instabilities are organically interactive within a complex world system that has reached crisis status at local, global, and planetary levels. Lying at the heart of this complex crisis are long-neglected disparities in health and well-being within and between countries, the refusal to face how these and climate change have arisen, and how economic considerations have fuelled the trend towards entropy (gradual decline of the planet into disorder). The critical point we have reached, starkly highlighted by the emergence of the COVID-19 pandemic *pari passu* with ongoing climate change and planetary degradation, reminds us of our global interconnectedness with each other and with nature. Comprehending and acknowledging the myriad, humanly constructed forces in each of these domains influencing all aspects of life, are the first steps towards effectively facing challenges to our health, our humanity (collectivity as humans) and our planet. Overcoming denial, acknowledging the magnitude and complexity of these challenges, prescient vision and dedicated action capable of fostering the cooperation for overcoming obstacles are now vital to seeking peaceful pathways towards more equitable and sustainable lives. South Africa is a microcosm of the world, with its local threats and challenges mirroring the global.

Significance:

Instabilities that pervade the world, highlighted by the COVID-19 pandemic, are especially significant for South Africa, where they manifest most starkly because of its apartheid legacy, its relative success economically on the African continent, and the implications of ongoing widening disparities and antagonism amongst South Africa's diverse people. Belief in moving towards narrowing wide disparities through decolonisation and reversion to an 'idyllic African heritage' via a transformation that includes widespread corruption, and the ANC government's perverse erosion of lives today and in the future through 'state capture', intensifies rather than ameliorates our predicament in an era when cooperation and a clear vision of current threats and future possibilities are desperately needed.

In an accompanying article, potential pathways towards a better future are offered through suggested shifts in paradigms of thought and action.

> *...our common goal should be 'healthy people in a healthy environment...part of an interrelated and interdependent community: the community of living things, the world of nature.*
>
> Russell Train[1]

Introduction

The COVID-19 pandemic is the latest of many newly emerging infectious diseases that have long been predicted in the context of escalating adverse impacts of human behaviour on our natural environment.[2] The pandemic's evolution in tandem with climate change and planetary degradation, exposes the increasingly dystopic state of the world, and highlights our moral and practical failures to put into action the detailed documented strategic plans that were developed after the SARS and Ebola warnings, to provide essential services to detect, report and respond to widespread outbreaks of infectious diseases.[3] These failures are due to insincerity in practising the values we claim as central to peaceful and sustainable advancement of human flourishing[4], as well as to the startling inability to take effective action on the even more ominous problem of climate change[5]. Conflicting and unsubstantiated messages in social media also play a role in obscuring the forces shaping our lives.[6]

Yet it was over 40 years ago that Johan Galtung warned of crises of violence, misery, poverty, the environment and repression of human rights, all rooted in *world structural violence* perpetrated through subtle, systemic, unequal distribution of power and resources.[7] Characteristics of our current humanly constructed global/planetary crisis include egregious disparities in wealth and health[8]; inadequate and inequitable healthcare services[9]; escalating frequency of emerging potentially fatal zoonotic infections[10,11]; food and water insecurity[12]; structural racism, displacement and refugeeism[13]; and ecological degradation[14]. All these perversions are deeply associated with a fraudulent global political economy's exploitative, accumulative and distributive mechanisms.[15,16] Resulting conflict, anger, and the potential for wars – civil, nuclear, cyber, cyborg or biological, obstruct or could even obliterate pathways towards peace and human flourishing.[17] For example, even a limited nuclear war between India and Pakistan, which between them possess only 1.5% of the global nuclear arsenal, could cause a nuclear 'winter' with up to 2 billion people starving to death.[18]

This complex crisis reveals the paradox of threats to the long-term health and survival of our species, notwithstanding all our technological advances made in science, health care, and the promises of genetic medicine, big data and artificial intelligence. Widespread failure to acknowledge and act constructively on long-standing insights into the fragility

of our planetary system, underlies the health challenges and ecological degradation that have fatal implications for life on our planet.[2,16,19,20]

Health

Advances in health and life expectancy

Impressive advances in science and medicine, and massive growth of the global economy have resulted in much-celebrated improvement in health during the past 100 years.[21] For example, life expectancy at birth has increased from under 40 years to over 80 years; maternal mortality rates fell from 385 per 100 000 live births in 1990 to 216 in 2015, smallpox and polio have been almost eradicated globally, deaths from cardiovascular disease have declined and many people now survive previously fatal malignant diseases.

Disparities in health

Despite these overall improvements, wide disparities in health remain, within and between countries.[8] For example, in Canada the lifetime risk of a woman dying from complications of pregnancy or childbirth is 1 in 8800 compared with 1 in 37 in sub-Saharan Africa.[8] Life expectancy at birth, ranging from 40 years to 80 years, is strongly related to absolute and relative levels of income. Within most countries, these patterns of difference also persist with similar, although narrower, differences in life expectancy and other key measures of health[22], as the visible components of a large 'iceberg' of the associated burden of disease[23] and social suffering that undermine so many lives[24].

Exacerbation of health inequities affecting people of colour and the poor in wealthy countries during the COVID pandemic, and in the poorest regions of the world (sub-Saharan Africa and Southeast Asia)[25], exemplify the contexts of crises linked to the socio-economic determinants of health, with deep discriminatory and exploitative roots. These inequities have fostered divisive, high-profile tensions within and between nations, and have highlighted the persistence of structural racism[26], once predominantly associated with overt racial apartheid in South Africa[27], but in reality a long-standing, inadequately acknowledged, much neglected feature of global racial and class discriminations[28]. Polarisation is aggravated by failure to see health as the social goal of making each of us safe only if all are safe by sharing vaccines equitably, avoiding relentlessly rising medical costs, widening disparities in access to health care and mistrust of science.[29]

Determinants of health

Health should be considered as a comprehensive 'state of being', profoundly influenced by complex interactions between our genetics/biology, our physical environment, economic and social systems, cultural and behavioural patterns, and scientific and technological advances in medical care. These interactive forces shaping health begin at conception, extend through pregnancy and childhood into adult life, and significantly determine longevity and how we die.

A public health perspective reveals that health is most powerfully determined by the social conditions of life (housing, sanitation, safe water, adequate nutrition and education), and the societal forces (political and economic ideologies) that profoundly shape living conditions.[30] This association was first noted in the ten-fold reduction in mortality from tuberculosis due to improved living conditions and use of sanitoria during the 19th century, *before* the development of effective drug treatment that could cure almost all patients with this disease.[31] More recently, it has been shown that even in a wealthy country like Canada, 50% of population health status can be explained by complex, intertwined socio-economic forces.[32] Access to modern biomedical health care accounts for 25%, biology and genetics for 15%, and the built and natural environment for 10%.[32] The proportions of these determinants of health vary widely within and between geographical boundaries. The social and societal determinants most powerfully affect Africa – a continent disadvantaged by widespread poverty and a legacy of exploitation that continues relentlessly through internal and external processes. COVID-19 case rates and death rates have been lower in countries whose citizens are willing and enabled to cooperate with social rules, for example, wearing masks and social distancing, than in countries lacking these characteristics.[33] This also has implications for compliance with vaccination for public health benefits.

Crises of health and health care

Health, and healthcare systems are shaped by knowledge applied through politically and ideologically driven power relations that vary with contingent historical and cultural influences at interpersonal, local, national, international, and global levels. In the United Kingdom and Europe after the Second World War, equitable access to innovative medical treatments were initially allied to the politically motivated creation of the welfare state with progressive taxation to provide, inter alia, health care, free at the point of delivery for all citizens. Over the next three decades much was achieved through these ideals and the development of socially supportive structures, mainly in the Organisation for Economic Cooperation and Development (OECD) countries.[34]

In low- and middle-income countries, the Primary Health Care Approach to health care in the 1950s and 1960s embraced a similarly egalitarian health policy. The emphasis on rural health centres, staffed by paramedical workers, achieved remarkable improvements in health in many countries at much lower cost. The strategy of 'Health for All' by the year 2000, approved by Member States of the World Health Organization (WHO) at the historic Alma-Ata conference, was to provide more equitable, appropriate, effective basic health care, and to address the underlying social, economic, and political causes of poor health.[34]

In the late 1970s and early 1980s, this trajectory of advancement was altered through neoliberal economic and political influences that shaped reform of healthcare services and financing during four subsequent decades. Healthcare priorities were increasingly driven by market values with emphasis on individualism, consumerism and competitiveness. This shift of emphasis away from such ethical principles as equity and the core values of social and economic development, towards so called *market efficiency*, that lacks ethical underpinnings, was part of a broader ideological shift towards political conservativism.[34] It is arguably this ideological perspective, with its associated economic policies, that lies at the heart of widening disparities in health and human well-being within an inadequately understood global health landscape shaped by political power structures interconnected with the politics of trade, economics, development, governance, and foreign and security policies.[35]

The structure and functions of healthcare services globally have thus been re-shaped over the past 50 years by a political dialectic between the conflicting ideas of health care as a socially valued service with equitable access to all, and health care as a *commodity* available to those who can pay.[34] The distribution of annual global expenditure on health care is illustrative. A minority of 19% of the world population accounts for over 85% of such expenditure, with annual per capita health expenditure ranging from USD33 in Somalia, to USD10 966 in the USA in 2019.[36]

All healthcare systems embedded within this structure share three shortcomings – in varying combinations and proportions across the world. (1) *Distortions* relate to inappropriately structured services to deal with local disease burdens. (2) *Dysfunctionality* arises from adverse commercial and bureaucratic intrusions on healthcare delivery. (3) *Unsustainability* results from fragmented services that are poorly coordinated and unbalanced because of failure to use explicit priority setting processes, to serve procedural justice in allocating resources when healthcare needs exceed the ability to supply these.[37] Deficiencies observed in implementing management and containment plans for pandemics have revealed the weaknesses of public health systems, and new lessons are being learned as pandemics unfold across the world.[38,39]

The question of universal and equitable access to health care – a goal widely stated as a global priority – needs to be examined in the above context. The WHO and others that are actively 'promoting' the rhetoric of universal access can only help make this possible by defining *achievable levels* of universal access, considering that about 70% of the world's population lives on less than USD10 a day, and 50% on less than USD4 a day.

Interacting global dysfunctions that contribute to our complex planetary crisis

Biological instability

Humans, animals, plants, and microbes live closely together, both symbiotically and antagonistically. The evolution of pathogens and emergence of new pandemics are influenced by human behaviour patterns that include ecologically destructive industries[40], wet markets and intensive animal farming with escalating use of antibiotics leading to resistance[10], excessive fishing, deforestation and widespread pollution of air, land and water, all within an increasingly complex human networking system[11].

Our understanding and use of molecular genetics that emerged from insights into the dynamics of microbial ecological and evolutionary interactions[41], together with antibiotics and vaccines gave rise to optimism by the 1960s that infectious diseases could be eliminated. Resurging pessimism with the HIV/Aids pandemic was intensified, both by the emergence of many other zoonoses and by rising antimicrobial resistance, to the point where some previously curable diseases, as exemplified by tuberculosis, would become very expensive to treat.[42]

The formulation of a convergence model for the emergence of infectious diseases[41] and Joshua Lederberg's call for a paradigm shift away from the aggressive 'war against microbes and disease' towards an ecological perspective on the inter-relationships between microbes, hosts and their environments through a 'germs eye view' carried a more expansive analogous health message regarding human interactions with nature.[43] Lederberg also presciently suggested that at the onset of the 21st century, infectious diseases were fated to remain a 'crucial research challenge of conceptual intricacy and global consequence', and that 'the future of humanity and microbes would likely unfold as episodes of a suspense thriller, that could be titled "*Our Wits Versus Their Genes*"'[43].

Warnings of new pandemics, as signs of a planet under unprecedented stress, encouraged the international community to devise detailed plans to increase global security in relation to epidemics as potentially major future threats to global health.[3] However, the selective nature of these policy initiatives ignored consideration of other closely intertwined system-based threats to global health and security.[44,45] COVID-19 has thus emerged and spread at a late stage in the trajectory of climate change, as one of many destructive outcomes of human behaviour. The role of laboratory manipulation of the virus remains unclear.[46]

Exceeding ecological limits

Accelerating industrial production and human consumption (of energy and material goods), has been underpinned most powerfully since the late 1970s by the relentless pursuit of profit and economic growth, regardless of the extent of environmental damage. The resulting climate change (rising temperatures and sea levels, storms, droughts, flooding, and wildfires); the pollution of air, land, seas and lakes; and the creation of new ecological niches, all adversely impact on sustainability of the biosphere.[47]

Between 1990 and 2015, carbon dioxide emissions rose by 60%, with the wealthiest 1% emitting more than double the total amount of the poorest 50%[48] who bear the brunt of environmental degradation and infectious diseases, e.g. the recent floods in Pakistan. The annual per capita carbon dioxide emission is 15.5 tons in the USA, 7.38 tons in China and 1.91 tons in India.[49] The USA with 4.2% of the world's population accounts for 25% of global carbon dioxide production, and China with 17.8% of world population for almost 30%. Traditional monitoring of carbon emissions by sector at national levels, with large trade traffic across widely disparate territorial regions, masks how production in one country drives impacts in another and alters the perspective on responsibility for emissions within a specific country by distorting the relative contributions of, for example, the USA and China.[50] The 7 million deaths in 2017 due to air pollution (16% of all annual deaths globally)[51] was three times the number of annual deaths from tuberculosis, malaria and HIV/Aids combined[52].

James Speth has documented the 50-year history of climate change denial and escalating investment in fossil fuel production driven by all US administrations from Jimmy Carter in the 1970s to Donald Trump.[53] Naomi Oreskes and Erik Conway have lucidly described how doubt was deviously created in the public mind about the adverse effects of climate change despite knowledge of the long-term implications.[54] Such revelations are sad testimonies to human avarice, stupidity and short-term perspectives on life.

Social disruptions

The extent of divergence in health and wealth in high-, middle-, and low-income groups, both within and between countries, is just one of many aspects of social instability and antagonisms. Other serious threats to social stability and security include ongoing exploitation and dispossession of the poor; inept governance, corruption and collusion with autocratic kleptocrats, inter alia within the weapons trade[55]; the deliberately structured transfer of material and human resources from the poor to the wealthy[56]; ineffectual development aid[57]; geo-political power imbalances[58]; erosion of democracy[59]; and interference with local development patterns in a range of countries by exploiting vast natural resources, many for military purposes, with little benefit to humanity[60].

Pervasive conflict, fear and anger, associated with atrocious living conditions and local wars, have contributed to displacement and mass migration of 80 million people in desperate attempts to seek better lives. Predicted increases to between 100 million and perhaps even 1 billion people by 2050[61] have almost unimaginable implications for all countries[62]. With half the world's population living in China, India, Southeast Asia and Indonesia combined, and another billion in Africa (whose coastal and desert regions are at most risk of climate change damage) excess mortality will predictably continue to be much greater in the Global South than in the Global North.[5] Such catastrophes are already exemplified by the conservative estimate of 50 000 deaths in 2000 attributable to climate change in the world's poorer and vulnerable populations.[63] Displacement and migration of desperate people from these ravaged areas will pose massive humanitarian challenges in the future, as predicted in an allegorical novel 50 years ago.[64]

Wars have killed millions and caused unspeakable suffering, as exemplified by the massive death tolls of the First and Second World Wars and by the recent reminder that

> *During the past 60 years, the United States has suffered a series of failed wars in Indochina, Central America, the Middle East, and Afghanistan….at mind-boggling human and financial costs …each of these wars produced mayhem and suffering, followed by an American retreat.*[65]

There will be no long-term winners if such local and international conflicts extend to nuclear war[18,67,68], bioterrorism[69], cyber wars[70], or cyborg wars[71] with their potentially devastating and irreversible effects.

The reality of these social instabilities undermines the idea of sustainable development, highlights the limited achievements of the many requirements of the Universal Declaration of Human Rights[72,73], and questions our failure to change the metaphor of 'progress' to '*developing sustainability*'[74].

Economic fragility

The emergence of many new infectious diseases and climate change, as well as instabilities in health care, social relations, biological and ecological domains, are deeply linked to an economic crisis described as a '*creatively destructive*'[75] '*market civilization*'[76], characterised by the pursuit of short-term profit, within an economic agenda driven most powerfully by the USA. With the economy as paramount, and dispossession of communities through enclosure of the commons and free riding on the environment, the self-centred frenetic pursuit of endless consumerism within a fraudulent economic system[15,16] erodes the moral soul of humanity while destroying our planet[77].

During the 20th century, the global economy expanded and became increasingly unstable. Under post-Second World War Keynesian

economic policies that enabled progressive taxation with high marginal tax rates, the annual global GDP increased from USD3.5 trillion in 1945 to USD18 trillion in 1980. Simultaneous encouragement of cooperation within civil societies, enabled the building of relatively egalitarian societies, with universal access to health care, education, and social security systems in some middle- to high-income countries.

The introduction of neoliberal economic policies in the late 1970s and early 1980s, promoted by Margaret Thatcher and Ronald Regan, unleashed privatisation, de-regulation of the market, aggressive competition, increasing freedom of corporations and financial institutions, radical reduction in marginal taxation and the creation of many new jobs, all with variable effects in different countries. The effects of this were noted in1992 by a former Wall Street executive:

> *For those who value justice and equality, the past 12 years have been a painfully wrenching time. We have endured twelve years of extreme conservatism of the American political system at its worst: government by the rich for the rich. Today the spirit of national community has been lost and the ugly divisions of race, sex and class are wider than at any time in recent history…. America today is a nation divided not only against itself but also against the world.*[78]

The continuing widespread dominance of neoliberal economic ideology driven by the aspirations of the economic elite, who re-regulated the market to suit themselves[79], increasingly destabilised the economy – especially since 2008. In this market-dominated world, economic growth and other benefits of 'progress' accrue mainly to a minority (20–25%) of the world's population, whose extravagant entitlements/consumption patterns damage our ecology and covertly devalue the lives of billions.

For another example, consider that from 2000 to 2015 the world had difficulty raising USD¾ trillion for the Millennial Development Goals. Yet during the 2008 economic recession, USD17 trillion of public resources (some of which could have been used for public benefit) was found in 3 months to compensate culpable banks and bankers. The loss of their homes by millions of Americans, and austerity measures applied to millions of others, represented socialisation of the losses, while the bank bailouts, including massive bonuses for bankers, represented privatisation of the gains.[80]

In the early 21st century, income and wealth disparities are illustrated by their distribution in 2011. In 2011, 71% of the world's population lived on <USD10/day, and 7% on >USD50/day.[81] The wealthiest 10% currently earn 50% of annual *global income* (and hold 85% of global wealth) while almost 4 billion people (50%) earn 10% of income (and hold about 1% of wealth).[82] Inequality has also continued to increase within countries, with many being denied even the very basic benefits of economic growth.[83] In relation to poverty alleviation, it is nonsensical to celebrate 1 billion people being 'lifted out of poverty' when the actual achievement between 2001 and 2011 amounted to daily per capita income increases from about 20–30 cents below USD2/day to about 20–30 cents above USD2/day.[81] These low incomes, and even per capita incomes up to USD5/day, are grossly inadequate to provide even rudimentary social services for minimally decent lives.

In the USA, President Trump's policies intensified the effects of neoliberalism, and reversed advances in economic and racial equality achieved through the New Deal in the era of civil rights. Life expectancy in the USA, which was similar to that in other high-income nations in 1980, has fallen 3.4 years below the 2018 average in the G7 countries. This is equivalent to 461 000 excess US deaths in 2018 alone.[84] An economically divisive market model based on a corrupt and fraudulent global political economy has flourished.[85] Deliberately imposed adverse structural forces such as unfair trade rules, and the sustainment of unpayable debts, engineered to some extent through devious means, widen rather than narrow local and global disparities[86-89] at enormous opportunity costs, while preventing meaningful reform beyond the existing paradigm[90].

The language of *war against disease* is commonly used to develop combative technological strategies through innovation and competition. This framing is extrapolated to global health, and buttressed by linking health to economic growth as the main indicator of progress.[74,91] Indeed, the term 'sustainable development' is premised on the idea that economic growth is the marker of development. However, material wealth, as measured by GDP, reflects the amount of money exchanged without reference to its social value. For example, increased expenditure on weapons and wars will increase the GDP to the same extent as similarly increased expenditure on health care. But the latter is of far greater societal value, illustrating that mere increments in GDP do not necessarily reflect 'progress'. Notably the unpaid, caring and home-building activities of women are not factored into GDP, despite their contributions to well-being and social cohesion.[92,93]

Conclusions

We live in an interdependent world, in which human well-being is affected by a fraudulent global political economy that has amplified wasteful consumption patterns of non-renewable resources, over-commercialised health care, and aggravated other social aberrations. Entropic changes in our biosphere are now reaching tipping points within an increasingly intricate and fragile global system[69,94] with the domino effects into the future not yet fully foreseen.

Although recognised for many decades, it has only been recently acknowledged that systemic, globally entrenched socio-political forces have led to a complex planetary crisis of unprecedented magnitude. Our long-term plight as a species, shockingly revealed by the COVID-19 pandemic's total disruption of all aspects of life globally, although differently for the wealthy and the poor, has highlighted differing forms of suffering: *economic* (disparities in wealth and opportunities), *social* (security, restriction of highly valued freedoms and access to health care and education), *racial* (respect for dignity), and *biomedical* (disease incidences and death rates).

Failure to understand the workings of a complex global system[16,39,95], particularly the pervasive destructive impacts of a wicked economic system on all the domains discussed, and failure to act appropriately, will likely result in our being consumed in the conflict, confusion, and chaos of a world in ongoing entropy[96]. With the COVID-19 pandemic, the writing is more clearly on the wall than ever before. Socio-political commitment to appropriate action is an essential immediate step towards improving health globally, followed by reducing the upstream forces that cause, sustain, and aggravate the impoverished lives of over half the world's population.

Some of the challenges lying ahead, the impediments to ameliorating these, and glimmers of hope for improving the future are outlined in the accompanying article.

Competing interests

I have no competing interests to declare.

Acknowledgements

I am grateful to the Editor and anonymous reviewers for constructive comments on the original submission. I am indebted to the University of Toronto for opportunities to meet and collaborate with many distinguished colleagues in other disciplines who have influenced my thinking during annual invitations from 2000 to 2019 to teach and continue with my research. Having a foot in both the Global North and South broadened my perspective on the challenges we face, as well as on the strengths and weaknesses across the globe in adapting to necessary change. Some of the data and analysis within these articles were included in the Stuart J Saunders Lecture (2019) and the Marinus van den Ende Lecture (2021) at UCT. I take this opportunity to acknowledge my wife, Evelyn, for her dedicated and unflagging lifelong support of my work.

References

1. Train R. 1969 APH address "Prescription for a planet," cited by Arras G. In: Planetary ethics: Russell Train and Richard Nixon at the Creation. Hastings Centre Report. 2020;50(3):23–24. https://doi.org/10.1002/hast.1127
2. Garret L. The coming plague: Newly emerging diseases in a world out of balance. New York: Farrar, Straus, & Giroux; 1994.
3. Smith MJ, Upshur REG. Ebola and lessons learned from moral failures. Public Health Ethics. 2015;8(3):305–318. https://doi.org/10.1093/phe/phv028
4. Benatar SR. Global justice and health: Re-examining our values. Bioethics. 2013;27(6):297–304. https://doi.org/10.1111/bioe.12033
5. Butler CD. Infectious disease emergence and global change: Thinking systemically in a shrinking world. Infect Dis Poverty. 2016;5(1), Art. #12. https://doi.org/10.1186/2049-9957-1-12
6. Treen K, Williams H, O' Neil S. How climate change misinformation spreads online. Carbon Brief. 26 June 2020. Available from: https://www.carbonbrief.org/guest-post-how-climate-change-misinformation-spreads-online/
7. Galtung J. The true worlds. New York: The Free Press; 1980.
8. Schrecker T, Labonte R. State of global health in a radically unequal world: Patterns and prospects. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 15–21. https://doi.org/10.1017/9781108692137.002
9. The Encyclopaedia of World Problems and Human Potential. Inadequate health services. No date. Available from: http://encyclopedia.uia.org/en/problem/136525
10. Benatar D. Animals, the environment and global health. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 304–315. https://doi.org/10.1017/9781108692137.025
11. Bloom D, Cadarette D. Strengthening the global response to infectious diseases in the 21st century, with a COVID-19 epilogue. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 51–75. https://doi.org/10.1017/9781108692137.004
12. Stockholm International Water Institute (SIWI). Water scarcity threatens food security for billions. 01 December 2020. Available from: https://www.siwi.org/latest/water-scarcity-threatens-food-security-for-billions/
13. Giovetti O. Seven of the worst humanitarian crises to know in 2022. Concern Worldwide US. 22 December 2021. Available from: https://www.concernusa.org/story/worlds-worst-humanitarian-crises
14. Butler C. Ecological ethics, planetary sustainability, and global health. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 281–292. https://doi.org/10.1017/9781108692137.023
15. Whyte D, Wiegratz J. editors. Neoliberalism and the moral economy of fraud. London: Routledge; 2017. https://doi.org/10.4324/9781315680545
16. Gill SR, Benatar SR. Reflections on the political economy of planetary health. Rev Int Polit Econ. 2020;27(1):167–190. https://doi.org/10.1080/09692290.2019.1607769
17. Mishra P. Age of anger. New York: Farrar, Straus, and Giroux; 2017.
18. Weston K. World War 3: India-Pakistan war could cause NUCLEAR WINTER – 'no one is safe'. Express. 27 February 2019. Available from: https://www.express.co.uk/news/world/1093062/World-War-3-india-pakistan-war-attack-video-nuclear-weapon-nuclear-winter
19. Garrett L. Betrayal of trust: The collapse of public health. New York: Hyperion; 2000.
20. Benatar SR. The HIV/AIDS pandemic: A sign of instability in a complex global system. J Med Philos. 2002;27(2):163–177. https://doi.org/10.1076/jmep.27.2.163.2992
21. Whitty CJM. Harveian Oration: Triumphs and challenges in a world shaped by medicine. Clin Med (Lond). 2017;17(6):537–544. https://doi.org/10.7861/clinmedicine.17-6-537
22. Benatar S, Sullivan T, Brown A. Why equity in health and in access to health care are elusive: Insights from Canada and South Africa. Glob Public Health. 2018;13(11):1533–1557. https://doi.org/10.1080/17441692.2017.1407813
23. Roser M, Ritchie H. Burden of disease. Our world in data [database on the Internet]. No date [cited 2022]. Available from: https://ourworldindata.org/burden-of-disease
24. Social suffering. Daedalus. 1996;125(1):1–283. https://www.jstor.org/stable/i20027347
25. Yaya S, Yeboah H, Handy C, Out A, Labonte R. Ethnic and racial disparities in COVID-19-related deaths: Counting the trees, hiding the forest. BMJ Glob Health. 2020;5, e002913. https://doi.org/10.1136/bmjgh-2020-002913
26. Jones CP. Levels of racism: A theoretic framework and a gardener's tale. Am J Public Health. 2000;90:1212–1215. https://doi.org/10.2105/AJPH.90.8.1212
27. Van Rensburg HCJ, Benatar S. The legacy of apartheid in health and health care. Afr Tydskr Sosiol. 1993;24(4):99–111. https://doi.org/10.1080/02580144.1993.10431680
28. Alexander T. Unravelling global apartheid. Cambridge, UK: Polity Press; 1996.
29. Oreskes N, Conway E. Why trust science? New York: Princeton University Press; 2019. https://doi.org/10.1515/9780691189932
30. Birn A-E, Kumar R. Societal determinants, and determination of health. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 28–50. https://doi.org/10.1017/9781108692137.003
31. Colgrove J. The McKeown thesis: A historical controversy and its enduring influence. Am J Public Health. 2002;92(5):725–729. https://doi.org/10.2105/AJPH.92.5.725
32. The Standing Senate Committee on Social Affairs, Science and Technology, Canada. A healthy, productive Canada: A determinant of health approach. Final report of The Subcommittee on Population Health [document on the Internet]. c2009 [cited 2022 Jan 30]. Available from: https://sencanada.ca/content/sen/Committee/402/popu/rep/rephealth1jun09-e.pdf
33. Gelfand MJ, Jackson JC, Pan X, Nau D, Pieper D, Denison, E, et al. The relationship between cultural tightness-looseness and COVID-19 cases and deaths: A global analysis. Lancet Planet Health. 2021;5(3):E135–E144. https://doi.org/10.1016/S2542-5196(20)30301-6
34. Benatar R, Sanders D, Gill S. The global politics of health care reform. In: McInnes C, Lee K, Younde J, editors. The Oxford handbook of global health politics. Oxford: Oxford University Press; 2020. p. 455–468.
35. McInnes C, Lee K, Younde J. editors. The Oxford handbook of global health politics. Oxford: Oxford University Press; 2020. https://doi.org/10.1093/oxfordhb/9780190456818.013.1
36. World Health Organization (WHO). Global spending on health: A world in transition. Report 2019 [document on the Internet]. c2019 [cited 2022 Jan 30]. Available from: https://www.who.int/publications/i/item/WHO-HIS-HGF-HFWorkingPaper-19.4
37. Benatar SR, Ashcroft R. International perspectives on resource allocation. In: Quah SR, Cockerham WC, editors. The international encyclopedia of public health. 2nd ed. Oxford: Academic Press; 2017. p. 316–321. https://doi.org/10.1016/B978-0-12-803678-5.00380-5
38. Garrett L. Ebola's lessons: How the WHO mishandled the crisis. Foreign Affairs. September/October 2015. Available from: https://www.foreignaffairs.com/articles/west-africa/2015-08-18/ebolas-lessons
39. Zacharia F. Ten lessons for a post-pandemic world. New York: W.W. Norton & Co.; 2020.
40. Schrecker T, Birn A-E, Aguilera M. Challenging the global extractive order: A global justice imperative. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 256–268. https://doi.org/10.1017/9781108692137.020
41. Institute of Medicine. Microbial threats to health: Emergence, detection and response. Washington DC: National Academy Press; 2001.
42. US Centers for Disease Control. CDC fact sheet: The costly burden of drug resistant TB disease in the U.S. [document on the Internet]. c2021 [cited 2022 Jan 30]. Available from: https://www.cdc.gov/nchhstp/newsroom/docs/factsheets/TB-in-the-US-508.pdf
43. Lederberg J. Infectious history. Science. 2000;288(5464):287–293. https://doi.org/10.1126/science.288.5464.287

44. Gill S, Benatar SR. Global health governance and global power: A critical commentary on *The Lancet* University of Oslo Commission Report. Int J Health Serv. 2016;46(2):346–365. https://doi.org/10.1177/0020731416631734

45. Benatar SR, Upshur R, Gill S. Understanding the relationship between ethics, neoliberalism and power as a step towards improving the health of people and our planet. Anthr Rev. 2018;5(2):155–176. https://doi.org/10.1177/2053019618760934

46. Van Helden J, Butler CD, Achaz G, Canard B, Casane D, Claverie J-M, et al. An appeal for an objective, open, and transparent scientific debate about the origin of SARS-CoV-2. Lancet. 2021;398(10309):1402–1404. https://doi.org/10.1016/S0140-6736(21)02019-5

47. Steffen W, Richardson K, Rockström J, Cornell SE, Fetzer L, Bennett EM, et al. Planetary boundaries: Guiding human development on a changing planet. Science. 2015;347(6223), Art. #1259855. https://doi.org/10.1126/science.1259855

48. Harvey F. World's richest 1% cause double CO_2 emissions of poorest 50%, says Oxfam. The Guardian. 21 September 2020. Available from: https://www.theguardian.com/environment/2020/sep/21/worlds-richest-1-cause-double-co2-emissions-of-poorest-50-says-oxfam

49. Worldometer [database on the Internet]. No date [cited 2022 Jan 30]. Available from: https://www.worldometers.info/co2-emissions/co2-emissions-per-capita/

50. Chen Z-M, Ohshita S, Lenzen M, Wiedmann T, Jiborn M, Chen B, et al. Consumption-based greenhouse gas emissions accounting with capital stock change highlights dynamics of fast-developing countries. Nat Commun. 2018;9:3581. https://doi.org/10.1038/s41467-018-05905-y

51. World Health Organization. Air pollution [webpage on the Internet]. No date [cited 2022 Jan 30]. Available from: https://www.who.int/data/gho/data/themes/air-pollution

52. Lancet Commission on Pollution and Health [webpage on the Internet]. c2017 [cited 2022 Jan 30]. Available from: https://www.thelancet.com/commissions/pollution-and-health

53. Speth JG. They knew: The US Federal Government's fifty-year role in causing the climate crisis. Cambridge, MA: MIT Press; 2021. https://doi.org/10.7551/mitpress/14083.001.0001

54. Oreskes N, Conway EM. Merchants of doubt. New York: Bloomsbury Press; 2010.

55. Feinstein A. The shadow world inside the global arms trade. Johannesburg: Jonathan Ball; 2011.

56. Caplan G. The betrayal of Africa. Toronto: House of Anansi Press; 2008.

57. Zwi A. Development assistance for health: Trends and challenges. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 207–220. https://doi.org/10.1017/9781108692137.016

58. Diamond J. Upheaval: How nations cope with crisis and change. London: Penguin House; 2019.

59. Levitsky S, Ziblatt D. How democracies die. New York: Penguin; 2018.

60. Sachs JD. America's confrontational foreign policy failed. It should pursue a cooperative global policy. Boston Globe. 03 September 2021. Available from: https://www.bostonglobe.com/2021/09/03/opinion/americas-confrontational-foreign-policy-failed-it-should-pursue-cooperative-global-policy/

61. UN High Commissioner for Refugees (UNCHR). Global trends: Forced displacement in 2019. Geneva: UNCHR; 2019. Available from: https://www.unhcr.org/globaltrends2019/

62. Straehle C. Mass migration and health in the Anthropocene epoch. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 293–303. https://doi.org/10.1017/9781108692137.024

63. Friel S. Butler C, McMichael A. Climate change and health: Risks and inequities. In: Benatar S, Brock G. Global health and global health ethics. Cambridge, UK: Cambridge University Press; 2021. p. 198–209. https://doi.org/10.1017/CBO9780511984792.018

64. Raspail J. The camp of the saints. Paris: Editions Robert Laffont; 1973.

65. Sachs JD. America's confrontational foreign policy failed. It should pursue a cooperative global policy. Boston Globe. 03 September 2021. Available from: https://www.bostonglobe.com/2021/09/03/opinion/americas-confrontational-foreign-policy-failed-it-should-pursue-cooperative-global-policy/

66. Marche S. The next US civil war is already here – we just refuse to see it. The Guardian. 04 January 2022. Available from: https://www.theguardian.com/world/2022/jan/04/next-us-civil-war-already-here-we-refuse-to-see-it?CMP=oth_b-aplnews_d-1

67. Solomon F, Marston RQ, editors. The medical implications of nuclear war. Washington DC: National Academy Press; 1986.

68. Kuzmarov J. What's behind the Biden administration's new $100 billion nuclear missile system? Covert Action Magazine. 9 March 2021. Available from: https://covertactionmagazine.com/2021/03/09/whats-behind-the-biden-administrations-new-100-billion-nuclear-missile-system/

69. Roffey R, Lantorp K, Tegnell A, Elgh F. Biological weapons and bioterrorism preparedness: Importance of public-health awareness and international cooperation. Clin Microbiol Infect. 2002;8:522–252. https://doi.org/10.1046/j.1469-0691.2002.00497.x

70. Sanger R. The perfect weapon: War, sabotage, and fear in the cyber age. New York: Crown; 2018.

71. Coyal M. Quantum leap: Intelligent robots & cyborg warriors will lead the charge in battle. The Economic Times. 29 December 2019. Available from: https://economictimes.indiatimes.com/news/defence/quantum-leap-intelligent-robots-cyborg-warriors-will-lead-the-charge-in-battle/articleshow/73011831.cms

72. Benatar SR. Annual Human Rights Lecture University of Alberta: Global health & human rights: Working with the 20th century legacy [webpage on the Internet]. c2012 [cited 2022 Jan 30]. Available from: https://www.ualberta.ca/global-education/visiting-lectureship-human-rights/past-lecturer

73. Forman L. International human rights law and the social determinants of health. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 122–135. https://doi.org/10.1017/9781108692137.009

74. Bensimon C, Benatar SR. Developing sustainability: A new metaphor for progress. Theor Med Bioeth. 2006;27:59–79. https://doi.org/10.1007/s11017-005-5754-1

75. Schumpeter JA. Capitalism, socialism, and democracy. London: Routledge; 1976.

76. Gill S. Globalization, market civilization, and disciplinary neoliberalism. Millennial J Int Stud. 1994;24(3):399–423. https://doi.org/10.1177/03058298950240030801

77. Robinson WI. Global capitalism and the crisis of humanity. New York: Cambridge University Press; 2014. https://doi.org/10.1017/CBO9781107590250

78. Driver DE. Defending the left: An individual's guide to fighting for social justice, individual rights, and the environment. Chicago, IL: Noble Press; 1992.

79. Cassidy R. 'Casino capitalism' and the financial crisis. Anthropol Today. 2009;25(4):2–13. https://doi.org/10.1111/j.1467-8322.2009.00676.x

80. Gill S, Bakker IC. The global crisis and global health. In: Benatar S, Brock G, editors. Global health and global health ethics. Cambridge, UK: Cambridge University Press; 2011. https://doi.org/10.1017/CBO9780511984792.020

81. Kochhar R. A global middle class is more promise than reality: From 2001 to 2011, nearly 700 million step out of poverty, but most only barely. Washington DC: Pew Research Center; 2015. Available from: http://www.pewglobal.org/2015/07/08/a-global-middle-class-is-more-promise-than-reality

82. United Nations. World social report. Geneva: United Nations; 2020. Available from: https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/02/World-Social-Report2020-FullReport.pdf

83. Atkinson AB. Inequality: What can be done. Cambridge, MA: Harvard University Press; 2015. https://doi.org/10.4159/9780674287013

84. Bor J, Himmelstein DU, Woolhandler S. Trump's policy failures have exacted a heavy toll on public health. Scientific American. 05 March 2021. Available from: https://www.scientificamerican.com/article/trumps-policy-failures-have-exacted-a-heavy-toll-on-public-health1/

85. Stiglitz J. The great divide. London: Allen Lane; 2015. https://doi.org/10.2175/193864715819556651

86. Elliott KA, editor. Corruption and the global economy. New York: Columbia University Press; 1997.

87. Mueller AP. What's behind the financial market crisis? Mises Daily. 18 September 2008. Available from: https://mises.org/library/whats-behind-financial-market-crisis

88. Gill S, Bakker IC, Wamsley D. Morbid symptoms, organic crises, and enclosures of the commons: Global health since the 2008 world economic crisis. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 242–255. https://doi.org/10.1017/9781108692137.019

89. Wiegratz J. The age of fraud: The link between capitalism and profiteering by deception. London School of Economics Blog. 13 November 2018. Available from: https://blogs.lse.ac.uk/africaatlse/2018/11/13/the-age-of-fraud-the-link-between-capitalism-and-profiteering-by-deception/

90. MccGwire M. The paradigm that lost its way. Int Aff. 2001;77(4);777–803. https://doi.org/10.1111/1468-2346.00219

91. Benatar S. Politics, power, poverty and global health: Systems and frames. Int J Health Policy Manag. 2016;5(10):599–604. https://doi.org/10.15171/ijhpm.2016.101

92. Bakker IC, Gill S. Power, production and social reproduction. New York: Palgrave MacMillan; 2003. https://doi.org/10.1057/9780230522404

93. Bakker I, Gill S, Wamsley D. Towards a new common sense: The need for new paradigms for global health beyond the COVID-19 emergency. In: Benatar S, Brock G, editors. Global health: Ethical challenge. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 470–477. https://doi.org/10.1017/9781108692137.040

94. Xu R, Yu P, Abramson MJ, Johnston FH, Samet JM, Bell ML, et al. Wildfires, global climate change, and human health. N Engl J Med. 2020;383(22):2173–2181. https://doi.org/10.1056/NEJMsr2028985

95. Elwell FW. Sociocultural systems: Principles of structure and change. Edmonton: AU Press; 2013.

96. Benatar S. A divided world in entropy. Society. 2018;55:200–206. https://doi.org/10.1007/s12115-018-0233-2

Note: A list of additional readings is provided in the Supplementary material of the accompanying article.

AUTHOR:
Solomon R. Benatar[1,2]

AFFILIATIONS:
[1]Emeritus Professor of Medicine and Senior Scholar, University of Cape Town, Cape Town, South Africa
[2]Adjunct Professor, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

CORRESPONDENCE TO:
Solomon Benatar

EMAIL:
Solomon.benatar@uct.ac.za

DATES:
Received: 30 Jan. 2022
Revised: 02 Aug. 2022
Accepted: 06 Oct. 2022
Published: 30 Nov. 2022

HOW TO CITE:
Benatar SR. Health in a post-COVID-19 world. S Afr J Sci. 2022;118(11/12), Art. #14995. https://doi.org/10.17159/sajs.2022/14995

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITOR:
Leslie Swartz

KEYWORDS:
new paradigms, post-COVID world, capacity for change, impediments, hope

FUNDING:
University of Cape Town

Health in a post-COVID-19 world

In the previous article in this issue (S Afr J Sci. 2022;118(11/12), Art. #13165), the emergence and spread of COVID-19 *pari passu* with climate change and planetary degradation were interpreted as late manifestations in the trend towards gradual decline into disorder (entropy) in an unstable and ecologically threatened planet. In this article, as we contemplate a post-COVID world, the question is whether new insights could generate courageous, prescient leadership towards new paradigms of health, politics, economics, society, and our relationship with nature. A gloomy prognosis is postulated because of the power of many impediments to such changes, both in an increasingly polarised world and in South Africa as a microcosm. Despite many squandered opportunities and a decline in local and global cooperation between all who have a stake in the future, some hope is retained for innovative shifts towards sustainable futures.

Significance:

Precarious local and global instabilities are vivid reminders of our interconnectedness with each other and with nature. Insights into local and global threats and opportunities, call for paradigm shifts in thinking about and taking action towards a potentially sustainable future in a country that has its own unique history and problems but is also a microcosm of the world. The impediments to making appropriately constructive paradigm shifts in many countries with their tendencies to authoritarianism that threaten peace and democracy, are even more complex in South Africa, where opportunities for dialogue and cooperation are diminishing. Retaining some hope, with vision and courage for innovative shifts towards a sustainable economic/ecological paradigm locally and globally, is arguably essential.

The key to good decision-making is not knowledge. It is understanding. We are swimming in the former. We are desperately lacking in the latter.

Malcom Gladwell[1]

Introduction

Our global predicament today can be viewed as an interregnum between returning us to the *unsustainable* status quo before the COVID pandemic, or to embracing paradigm shifts that could propel human beings collectively towards developing *sustainable* long-term goals. The short-sighted choice would be limited to the current pandemic control measures, new vaccines, some strengthening of healthcare systems, small-scale poverty alleviation, and development of renewable forms of energy. A more desirable trajectory would include a new vision of the future, supported by transdisciplinary teams of scholars/researchers evaluating means of changing complex global systems, and public participation in creating the political will and administrative capacity to grasp such daunting challenges.[2,3] Wise stewardship into the future would be characterised by more modest material entitlements, enhanced attention to energy conservation and making peace rather than war.

Which path will be chosen? By whom will the choice be made? Will we continue as a global community along the current socially, ethically and ecologically unsustainable trajectory of 'progress', defined as expanding knowledge, developing new technology, promoting endless economic growth and philanthropy? Are we trapped within the ideological illusion of endless, market-driven, economic growth that currently pervades and corrupts all aspects of our lives? Alternatively, do we have the will and capacity to take the more difficult, but potentially sustainable trajectory through new framings, metaphors, and paradigms of living[4-7], with a new ethics for 'being' that is sensitive to future generations?[8-11] These ideas, expressed briefly here are more expansively articulated in many books and articles, some referenced herein with many more cited in those publications and in a rapidly expanding literature.

Return to the status quo ante?

It is unlikely that the world could be returned to a resemblance of the status quo ante. Yet it seems from current trends that the most privileged and powerful, with their conservative tendencies, are attempting, as in the past, to restore *their normality* without reducing disparities or their consumption patterns contributing to climate change. Within this scenario, climate change disasters would escalate on a continually degrading planet[12,13], with newly emerging infectious diseases and increasing anti-microbial resistance. Thirty years ago, US Vice President, Al Gore, opined that those with '…a vested interest in the status quo will probably continue to stifle any meaningful change, until enough citizens are willing to speak out and urge their leaders to bring the earth back into balance'[14].

This path further reveals the current global economic system as both fraudulent and immoral – even considered as a crime against humanity[15,16] for having spawned both extremely wealthy and desperately poor population groups through its focus on profit and wealth accumulation by exploitation of the middle and poor classes, and free-riding on the environment[10,11,17]. The resulting perpetuation of divergencies in health, wealth and living status, exemplified by life expectancies at birth that differ by up to 30 years, the unseen suffering associated with egregious differences in wealth and health, and neglect of our vital biosphere, could be interpreted as consequences of ecocidal activities with genocidal effects.

In her analytic comparison of the world in 1914 and 2014, Margaret McMillan perceived disquieting parallels between failure to understand the factors that propelled Europe to war in 1914, and events in the world 100 years later that have placed us at risk of a similar catastrophe.[18] Paul Rogers concurred that by 2020, available evidence inspired little confidence that there was sufficient wisdom to avert the destructive potential of some scientific and

technological developments, even if inevitable catastrophes had not yet materialised.[19] John Gray contended that states' struggle for power over natural resources, had led to an era of geopolitical rivalry reminiscent of a century ago, but with new participants and higher stakes.[20] Other eloquent descriptions of the ravages of the global political economy, reveal the extent to which our inability to learn from history contributes to failure of the much cherished projects of international peace and world government.[21-23] Consequent ongoing human conflict and domino effects will enhance a trajectory of progressive ecosystem entropy, with a potentially fatal outcome for human life on our planet.

A new trajectory of progress: Changing our global/planetary paradigm

Recognition and acknowledgement of tipping points

Our position at many tipping points[24,25] makes it essential to acknowledge that solutions to 21st-century problems of global health and security are not accessible solely through scientific and technological innovation, biomedicine and endless economic growth within our current competitive, hyper-individualistic paradigm that has given rise to these problems[26,27]. *Knowing how* ideas sparked the 'swerve' into the modern age through the Renaissance[28] could provide the impetus to pursuing another 'swerve' through new paradigms of health, politics, economics and social relations towards improved lives on a sustainable planet[29]. In contemporary language, this would involve updating our cerebral software and its now defunct 'processes' with a new program geared to the reality of our time.

Changing the health paradigm

The first challenge is to understand that dangerous health inequities within and between countries cannot be addressed by highly technical, individualised approaches to health care. Contemplating health from the broadest perspective is crucial to revealing the myriad forces influencing population health that need to be corrected. The valid criticism that a narrow *focus* on care predominantly for patients with COVID excessively displaced care for patients with other often urgent medical and surgical needs, does not require discarding the biomedical focus on diseases that is so vital to caring for individual people through medicine's enduring bedside technical and humanitarian skills.[30] It does mean that a comprehensive way of thinking about population health is needed by extending the notion of health beyond the individualistic, biomedical model towards a holistic, ecological perspective.[2,5,9]

In the face of still evolving global economic and health crises in a low-growth economy, a vital question for upper- and middle-income countries is not whether more money is needed for health care or the well-being of their citizens, but rather whether available resources could be spent more wisely. The goal would be to protect/improve social infrastructures and provide adequate social living conditions with optimal, easily accessible and effective care services. This could be achieved through transparent, accountable resource allocation procedures[31], and a continuing quest for greater equity in balanced healthcare expenditure on individual health and public health through coordinated, evenly spread teams of health professionals[32-34].

Changing the political paradigm

At the beginning of this century, Michael MccGwire argued that the adversarial national security paradigm's evolution over the previous 60 years, shaped by beliefs of specific times and events, had lost its way.[35] He described this now dying paradigm as characterised by: (1) *relationships* of exclusion, confrontation, domination and enmity; (2) *diplomacy* that is adversarial, intransigent, unilateral, vengeful and exploitative; (3) *power* that maintains superiority through compulsion and punishment; and (4) *security* based on inequality, deterrence, coercion, and national interests. He then advocated a shift to a new paradigm characterised by: (1) inclusive and engaging relationships; (2) cooperative and reciprocal diplomacy; (3) the use of power to persuade and reward; and (4) the pursuit of security through re-assurance and cooperation on a global scale.[36] Much remains to be learned and implemented from his prescient analysis and from new political ideas for an era beyond the Anthropocene.[37]

Changing the economic paradigm and the values guiding global governance

Examining the impact of the political economy on planetary health provides deeper insights into our current predicament and highlights the need to strive for a peaceful, globally sustainable economy.[38,39] In so doing we should contemplate that during the 4½ billion years' history of our planet, *Homo sapiens* appeared about 200 000 years ago and 'modern' life began only 10 000 years ago with the Agricultural Revolution. Carbon dioxide levels in the atmosphere remained somewhat stable at below 300 parts per million, and life on earth was sustainable until the First and Second Industrial Revolutions 270 years ago, and were intensified by the first oil-producing wells in 1857. Fossil fuels accumulated over millions of years have been exponentially extracted and consumed, especially over the past 50 years. Increasing easy access to food and improved living conditions enhanced the quality of human lives, extended life expectancy at birth from less than 40 years to almost 80 years for some and allowed the world population to increase from 1 billion people in 1800 to 2.5 billion in 1950 and almost 8 billion today.

Widening disparities in health and wealth, despite massive economic growth and global ethical frameworks such as the Universal Declaration of Human Rights, questions the values that direct our actions today in global health. Global governance for health[40] has been critiqued on grounds inclusive of duty dumping[41], exploitation[42] and incoherence[43,44]. The foundations of modern society's current approaches, based on economic liberalism, corporatism, managerialism, a focus on biomedical science, respect for human rights limited to civil and political rights, and healthcare services within an exploitative 'medical industrial complex', are distorted by power asymmetry and so-called 'global social norms' that limit the range of choice and constrain corrective action on health inequity.[40,42,45]

J.K. Galbraith's insights into the complacency of affluence and the need for a humane economic agenda[46,47], as well as many cautions against being fooled by a window-dressing agenda in feel-good societies, remain relevant. Ignoring such advice while conditions of life remain desperate for the majority in the face of continuing exponential consumption of energy by a minority, without concern for the future, augurs poorly for achieving a secure world. Global challenges and crises will likely get worse within such a trajectory, as presciently predicted indirectly decades ago[48,49] and directly more recently[50,51]. Such trends reveal limited insight into the health implications for wealthy countries if an uncontrolled pandemic becomes prolonged[52], and have neither moral justification nor any basis in any ethical theory. Reports that 50% of the initial COVID-19 vaccine supply was purchased for use by 13% of the world population, reflects the striking lack of solidarity with distant others (vaccine nationalism) during a devastating global pandemic.[53]

Correction requires new mental software capable of seeking global political solutions to counter powerful global political forces detrimental to health and planetary sustainability. The challenge is to develop an ethically justifiable economic path towards solidarity, cooperation, interdependence, a sense of responsibility for future life on earth, a fairer international distribution of resources[31], and a continuing quest for greater equity in health care.

A good starting point would be to redefine severe poverty more realistically, from the current level of per capita income of less than USD2/day to USD4–5/day, to catalyse significant poverty alleviation. This could strongly supplement poverty alleviation endeavours such as the Brazilian Bolsa Familia scheme and others cited in the WHO Commission on the Social Determinants of Health report.[54]

Sustainable health needs to be rooted in commitments to global solidarity and shared responsibility, and in wise global governance for health within an economic and political system that serves a global community of healthy people on a healthy planet. Rectifying the regnant economic system must be preceded by recognition of how this is sustained by distortions of our values; for example, the claimed rights by some for excessive freedom to satisfy endless entitlements without attention to the associated balancing responsibilities that are essential to satisfying rights and protecting the common good.[26]

A comprehensive reconsideration of the current economic paradigm[55] should be enhanced by questioning the validity of excessively affluent lifestyles. More resources should be allocated to improving the public infrastructure and institutions of social reproduction (e.g. education, health care), and placing quantifiable monetary value on the unpaid work of women for contributions to society that are currently not included in measurements of GDP. Revisions to how debt is created and perpetuated, as well as new forms of taxation, would be vital to supporting better lives and health in low- and middle-income countries, and to discouraging wealthy countries from living at the expense of future generations.[56]

Social innovation

Enlightenment values of freedom, rationality and economic growth are the drivers of innovation. Yet innovation is preferentially limited to siloed scientific progress and improvements in technological applications, not least in the predatory power of war to accumulate resources as a mechanism for survival.[57] New broadly based socially innovative ideas and projects are needed. For example, a 'not-for-profit world beyond capitalism and economic growth'[58], aimed at narrowing egregious disparities in human flourishing, could shift the global mindset by influencing the hearts and minds of whole populations. Collective social efforts to reduce anger and conflict through wider community interest and stake-holding in developing a sustainable future, shaped by innovation, political commitment, and social action[59,60], could become a means of promoting a shift towards solidarity and cooperation within a new narrative for planetary survival in a potential Symbiocene era[61,62].

Expanding the bio-political-ethical discourse

In 1971,Van Renselaar Potter argued for the integration of biology, ecology, medicine, and human values within the concept of a new term 'bioethics'.[63] Those who were developing their own conception of biomedical ethics largely ignored Potter's prescient warning about unsustainable progress, particularly in healthcare systems. The moral compass they developed was almost entirely focused on individuals and protection of civil and political rights with little attention to the public good.

In the 1990s, Hans Kung's pursuit of the idea that antagonistic ways of thinking and acting would not be sufficient to resolve 21st-century challenges, amazingly succeeded in getting all the religions to sign a doctrine of mutually supported beliefs.[64]

Subsequently, secularly based recommendations (and the rationale for these) were made to extend the bioethics discourse beyond the ethics of interpersonal relationships to include the ethics of institutional and international relationships, and ecological ethics in the pursuit of public, global and planetary health.[4,8,9] This was followed by a critique of the global political economy[10], and an interphilosophies dialogue methodology to facilitate peaceful communication across divisive cultural beliefs[65].

The deeply disturbing ethical insensitivity and inadequate responses to political threats of mass human violence, including nuclear weapon proliferation[66], more wars and major pandemics that threaten to annihilate and disrupt many lives, point to the urgency for stimulating our moral imagination regarding the human predicament[4,67,68].

Changing our attitude to nature

Extensive use and depletion of non-renewable natural resources, with consequent global warming, rising sea levels, floods, fires, and degradation of land, sea, and air, with multiple interlinking adverse effects on health[11,13,21,39] make it critical to radically reduce ecology-damaging use of energy[69]. In 2000, the World Bank's vice-president for Europe identified 20 specific problems regarding the global environment that required action within 20 years.[70] These goals were subsumed under three headings: (1) sharing our planetary concerns regarding the global commons; (2) sharing our humanity through issues requiring a global commitment; and (3) sharing a rule book for issues needing a global regulatory approach. Little attention has been paid to these recommendations aimed at changing a complex global system, and his goals have not been remotely achieved. These failures are exemplified by the increasing use of coal by 17% in the US in 2020, after a slow decline in the preceding years, and persistent refusal by successive US Congresses to support climate change legislation.[71]

Much could be done by each of us individually as well as by nations collectively. Ecology-preserving lessons could also be learned from indigenous people's stewardship of nature. Energy conservation through a range of feasible greening activities becomes essential. These would include dietary shifts away from animal products to plant foods; reduction in domestic and business energy use for lighting, cooling in summer, heating in winter, as well as increased use of solar power, reduced wasteful night-lighting in cities, unnecessary global travel and enhanced use of public transport. Lower levels of air pollution, and renewal of degraded natural resources, evident from reduced global travel during the COVID-19 pandemic, are supportive evidence for such suggestions.

It is incumbent on all of us globally to share in reducing both energy use and our ecological footprint to sustainable levels. Fairness requires that the greatest onus should fall on those with the highest per capita ecological footprint. Some recent steps in the right direction to address such issues of profound importance are gratifying but come late considering earlier insights[72], and remain associated with high levels of ongoing denial[73], duplicitous government behaviour[74], and obstruction to relevant legislation[71]. The already well-advanced development of alternative sources of renewable energy must continue, although it is unlikely that enough could be provided in time to avert ongoing climate change and planetary degradation[75] before many coastal cities and islands have become submerged, costing many lives and making the existence of survivors more precarious[76].

Are we capable of making radical global changes?

Prerequisites for making potentially more effective changes include widespread social acceptance of the need to ensure our future survival through prescient and effective global political leadership and health governance.[77] The question is whether we can muster sufficient solidarity through political[38,39], public[3], social[78], moral[6,8,79], and ecological[11] imaginations to find new solutions to egregious disparities[80]. These utopian paradigm shifts seem unlikely in the face of many impediments.

Impediments to change

While there are many reasons for denial, resistance and inaction, it is plausible to suggest that a dominant barrier to meaningful change lies in reluctance – even refusal – to admit that our predicament is in large part attributable to privileged populations' (wastefully) consuming highly disproportionate levels of energy. An inordinate sense of entitlement, and no real desire to reduce their consumption is reflected in the myopic view stated by George H.W. Bush at the first Earth Summit in Rio de Janeiro in 1992 that 'The American way of life is not up for negotiation. Period.'[81] This attitude does not seem to have changed since then. It has also been buttressed by the explanation that we lack the evolution of 'hard wiring' to face future threats![82] The implausibility of this excuse is that it offers a narrow biological explanation rather than a broader social one. Our ability to face threats in the future *is evident* in the extent to which we purchase life and other insurance policies, and even more strikingly in how we invest in a massive military machinery in anticipation of future security threats.

The major impediment to *shifting from paradigms* of living that have resulted in such a small proportion of people in the world being privileged (about 20%), is the invisibility of power structures and a belief system (in which power is embedded and perpetuated) that determines the dominant way of thinking and the framing of ideas, values and actions [see Note[a]]. Daniel Kahneman explains our predicament by suggesting that 'We can't live in a state of perpetual doubt, so we make up the best story possible and we live as if the story were true.'[83] Currently dominant belief systems, their frames and metaphors for global thinking are characterised by an emphasis on individualism, freedom and a market economy that places economic considerations above all else.

Our global plight is not entirely due to failure to pursue such hard-won and highly prized values as individual rights, tolerance, self-

determination, and democracy with checks and balances on political power, that have contributed to advancing human well-being. We are also culpable of pursuing distortions of these values in ways that improve the lives of some while excluding others who constitute the vast majority. For example, a self-destructive mode of life is propagated when individualism becomes hyper-individualism and when freedom of the powerful reduces the freedom of the weak. When human rights discourse is narrowly focused on civil and political rights – with little attention to social, cultural, and economic rights[28] – short-term interests are highlighted within what has become the favoured minimalist moral compass in secular societies[84]. Fraudulent economic dogma that pervades all aspects of life further dilutes other values such as a sense of community and solidarity with others. Additional examples include ignoring the interdependence of all life within a natural world of limited resources, toleration of a medical research agenda that is heavily skewed towards illnesses afflicting the wealthy, and marketing new therapies at large multiples of the cost of production.

The Lancet/University of Oslo Commission Report[40] illustrated the failure to acknowledge the multifactorial sociological underpinnings of ongoing global crises, and the role of the commitment of the wealthiest to fraudulent economic policies in the complex causal chain of multiple threats to global health and security.[85] While the Commission made an accurate diagnosis of global health inequities, the solutions proposed were closely linked to the structures and processes that caused the problems, and failed to link into the structures of both the governance for global health and the global political economy as some of their root causes.[42,43]

Most of the world's poor are people of colour, and the roots of structural racism and structural poverty are complex.[86-88] Honesty requires that the long-overdue attention now being paid to marginalised communities in the Global North should be extended to acknowledgement that impoverished lives in Africa and in the rest of the Global South also matter.[89] All lives are of value, regardless of wealth or educational level, and we are all diminished and threatened by allowing billions to live with preventable and unnecessary suffering.

Other overlapping rationalisations for inadequately addressing these human imperfections include a trajectory of history too complex and opaque to be widely understood and accepted[90]; the extreme discomfort of confronting human failures; the depth of shame and regret aroused by structurally discriminatory forces locally and globally[91]; the complexity of engaging open-mindedly in dialogues between contrasting belief systems[92,93]; distortions of our value systems; and lack of visionary leadership. There is also a strong tendency to avoid contemplating paradigm shifts in thinking and to restricting the expectations of the most privileged who mortgage the lives of future generations.[94] Confronting power relationships[95] and structural violence[96] that support deeply entrenched cultural patterns with their revered narratives, are further aggravated by global geo-political antagonisms with pursuit of power and wealth that preclude visionary collaboration and action across many deep cultural and socio-political-economic barriers[97-99].

John Maynard Keynes observed that 'the difficulty lies not so much in developing new ideas, as in escaping from old ones'[100]. The big question is: Can our species undertake and achieve the ambitious goals outlined here? Given the relentless pursuit of power and resources to meet unsustainable lifestyles, together with denial of human culpability, it seems highly unlikely that we have the volition or ability as a global community to escape from a competitive and destructive paradigm and move into a globally cooperative future. Friedrich Hegel expressed such skepticism and he warned that 'What experience and history teach is this – that nations and governments have never learned anything from history or acted upon any lessons they might have drawn from it.'[101]

Glimmers of hope

There are several reasons for retaining glimmers of hope. The first relates to our having acquired some insight into the fact that ongoing 'progress' is not sustainable and that both human advancement and social retrogression are becoming inescapable features of life. Our human capacity for empathy, as witnessed so frequently in everyday life, and the potential for some sense of global citizenship and solidarity under the threat of possible extinction, add hope to our ability to shape the future. Recent focused attention on the serious shortcomings of neoliberalism[102] and on imagining better worlds[3,4] is also gratifying.

Our ingenuity in addressing such complex issues as the human genome and targeted medical therapies, as well as space exploration and travel, suggest that we also have the ingenuity to address the complexity of social innovation challenges on a grand scale. The magnitude of the financial investment required to pursue such work is within our reach as mentioned in relation to the 2008 financial crash[10], and given that today the International Monetary Fund provides subsidies of USD5 trillion each year to the fossil fuel industry[103], and that almost USD2 trillion is spent on the military globally each year. If we could resolve to use our intellectual and material resources to make peace instead of war, and use our human ingenuity to undertake research and education on socially innovative changes, impressive advances could be achieved. While transition towards renewable energy will not be sufficient in the timeframe available, devoting more attention to reducing energy consumption, and doing better with less could be of great value if these were to become high priorities.[104]

The polarised extremes of several systems of belief (world views), each with their own ontology (what it means to be human), epistemology (how knowledge is defined) and axiology (study of the nature of values and valuation), provide conflicting challenges in many aspects of life. It has become popular to assert that the world's dominant belief system, caricatured as western, white male and capital-oriented, lies at the heart of the problem, and that solutions lie in replacing this with one or other version of indigenous and identity-related 'decolonised' ideologies. It is vital to acknowledge the roles of empires, colonisation and other means of creating 'the other' that have oppressed so many and undermined their health and dignity.[105] However, what decolonisation and transformation entails and how these would be implemented remain vague (at least in South Africa).[106,107] But it is surely evident that, rather than reciprocated revengeful attitudes, it would be more constructive to embark on mutually respectful, innovative, cooperative dialogue to identify overlapping values within the best notions of both the western and alternative world views. Humility will be needed to implement well-described and justifiable effective methods from other world views to deal with health issues responsive to such measures, for example, as argued in relation to the mental health of culturally abused children of residential schools in Canada, when western methods of psychotherapy are ineffective or even harmful.[108]

Such collaboration should be possible without rejecting or diminishing the value of reason and the scientific method. Emphasis is also needed on enhancing relationships within broader spectra of world views than those caricatured by their polar extremes. Full participation in the dialogue by all parties is essential to use the tension between competing ideas creatively rather than destructively. Ethical motives and reasoning can assist in recognising that while relationships within families and communities are primary, solidarity within an extended web of relations should also be highly valued. Indeed, components of both the dominant tradition and of alternative philosophies remind us of the greater complexity of real-world relationships, and direct us towards recognition and support for essential ethical relationships within and between communities, institutions, and nations, and with our natural environment.

Jared Diamond has opined that we can learn from the errors that led to the collapse of previous civilisations.[109] Bill Gates is also optimistic and centres his hope on technological innovation that could include geo-engineering.[110] Martin Rees[111], an imaginative scientist, has expressed faith (probably overinflated) in the benefits of artificial intelligence and space exploration for the future of humankind. Niall Ferguson gingerly shares some of this optimism but with the following caveat:

> *It would be wonderful if the advance of science had liberated us from at least some of the irrational ways of thinking that characterized the ancient and medieval world… But other ways of magical thinking have grown even as religious belief has diminished.*[112]

Vaclav Smil, who has intensively focused on using the best knowledge available to advance the human predicament, is neither entirely skeptical that we are facing an apocalypse nor dismissive of new pathways through scientific advances towards long-term peaceful survival.[82] David Graeber and David Wengrow in their monumental reinterpretation of human history, with emphasis on the unacknowledged contributions of indigenous societies, have also offered new positive insights into the forces that could shape a better future.[113]

Power dynamics are critical. Rather than concentrating on *power over others* (money and force), *power with others* (community values) has the potential to allow us to forge collective goals that have become vital to survival. This alternative perspective envisages a *shift in metaphor from competition to cooperation*. Judicious use of financial power, the power of empirically based knowledge and the power of well-reasoned moral authority are the three forms of power that could be used non-coercively to reduce the social instability that results from preventable poverty and disease that pose major security threats to the health and lives of all across the globe.

Conclusions

Framing global health in terms of a healthy planet requires an understanding of the adverse impact of our transition from a species dependent on nature for our survival, to one capable of manipulating nature, and most recently to ambitious (even overambitious and potentially harmful) endeavours to re-engineer life and the biosphere.[114]

Any serious attempt at improving global health will have to address such tough questions as: Are we willing to acknowledge what lies at the core of global health inequities? Can we promote deeper insights into who we are? Can we revise our expectations as global citizens and internalise the extent and nature of our responsibilities to those who are distant or less fortunate? Can we restore a balance in undistorted enlightenment values and ways of thinking? Can we communicate with those who have different belief systems, and in so doing craft an improved shared belief system? Are there visionary local and global leaders willing to face these challenges?

These questions could also be asked about South Africa, a country that is a microcosm of the world. The sorry state of its failed aspirations and the new government's shameless, corrupt use of power for the self-interest of its leaders, their friends and families rather than the population it should serve, have been eloquently revealed.[115-117] With the global being visible within the local, and vice versa, the message of interdependence is highlighted. Cooperation and building on the best are more likely to offer optimism than reactive destruction in the hope of building anew.

Ongoing advances in science and technology will assist in making progress towards a better world, but it has become essential to address the adverse upstream economic and social forces that are radically reshaping the viability of our planet and all its inhabitants. For hope to be retained it is essential that human ingenuity and benevolence will be applied to social innovation to promote peace, reduce inequities, and enhance democracy nationally and internationally.[118] With the evolution of the COVID pandemic and the ongoing emergence of more infectious diseases such as monkeypox, the writing is more clearly on the wall than ever before.

Pursuit of the ambitious trajectory described above is *the* crucial challenge facing humanity at a profoundly threatening time in the history of life on our planet. Rather than being daunted by the enormity of the task, hope should be garnered from the example of imaginative planetary systems-thinking, sustained research, collaborative strategies, and success in raising the resources for exploration of outer space.

The quest for sustainable global health will be elusive if we continue to neglect the upstream forces, particularly the pervasive destructive impacts of a wicked economic system on all the domains discussed that cause, sustain, and aggravate the impoverished lives of over half the world's population. Failure to act, will likely result in our being consumed by the conflict, confusion, and chaos of a world out of balance and in entropy. Deriving dazzling pleasure from seeing millions of light years into the past should not blind us to what lies ominously in front of our eyes! Although the outlook for constructive change is bleak, the ideas briefly outlined here could hopefully be used as study springboards to stimulate such work. Promoting a sense of 'global consciousness' within educational systems to enable children, adolescents and young adults to identify more effectively with others locally and globally, should be included in developing character and values for the 21st century. Educational institutions at all levels should include these goals as a central component of their endeavours.

[a]***Note:*** *Frames and metaphors are mechanisms of mind that shape our perceptions and structure our most basic understandings of our experience and actions that are used 'unconsciously', automatically and in repetitious ways to determine how knowledge is constructed and debated.*[119] *They allow us to understand reality, to create what we take to be reality and to facilitate our most basic interactions with the world by structuring our ideas and concepts, shaping the way we reason, and impacting on how we perceive and how we act.*[6] *The dominant belief system, as with other contrasting belief systems, mobilises feelings and motivations through symbols that work most powerfully when subconscious. What is believed becomes an important aspect of reality, whether true or not, and this applies to religious and secular belief systems.*[120]

Competing interests

I have no competing interests to declare.

Acknowledgements

I am grateful to the Editor and anonymous reviewers for constructive comments on the original submission. I am indebted to the University of Toronto for opportunities to meet and collaborate with many distinguished colleagues in other disciplines who have influenced my thinking during annual invitations from 2000 to 2019 to teach and continue with my research. Having a foot in both the Global North and South broadened my perspective on the challenges we face, as well as on the strengths and weaknesses across the globe in adapting to necessary change. Some of the data and analysis within these articles were included in the Stuart J Saunders Lecture (2019) and the Marinus van den Ende Lecture (2021) at UCT. I take this opportunity to acknowledge my wife, Evelyn, for her dedicated and unflagging lifelong support of my work.

References

1. Gladwell M. Blink: The power of thinking without thinking. New York: Little, Brown & Company; 2007.
2. Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. https://doi.org/10.1017/9781108692137
3. Faessel V, Falk R, Curtin M, editors. On public imagination: A political and ethical imperative. New York: Routledge; 2019. https://doi.org/10.4324/9780429343599
4. Benatar SR. Moral imagination: The missing component in global health. PLoS Med. 2005;2(12), e400. https://doi.org/10.1371/journal.pmed.0020400
5. Benatar SR. Health: Global. In: Ten Have H, editor. Encyclopedia of global bioethics. Cham: Springer; 2016. https://doi.org/10.1007/978-3-319-05544-2
6. Benatar SR. Politics, power, poverty and global health: Systems and frames. Int J Health Policy Manag. 2016;5(10):599–604. https://doi.org/10.15171/ijhpm.2016.101
7. Gray B. How should we respond to the global pandemic: The need for cultural change. Challenges. 2021;12(1), Art. #8. https://doi.org/10.3390/challe12010008
8. Benatar S, Daar A, Singer P. Global health ethics: A rationale for mutual caring. Int Affairs. 2003;79:107–138. https://doi.org/10.1111/1468-2346.00298
9. Benatar SR, Gill S, Bakker I. Making progress in global health: The need for a new paradigm. Int Affairs. 2009;85:347–371. https://doi.org/10.1111/j.1468-2346.2009.00797.x
10. Gill S, Bakker IC. The global crisis and global health. In: Benatar S, Brock G, editors. Global health and global health ethics. Cambridge, UK: Cambridge University Press; 2011. https://doi.org/10.1017/CBO9780511984792.020

11. Butler C. Ecological ethics, planetary sustainability, and global health. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 281–292. https://doi.org/10.1017/9781108692137.023

12. Xu C, Kohler TA, Lenton TM, Svenning JC, Scheffer M. Future of the human climate niche. Proc Natl Acad Sci USA. 2020;117(21):11350–11355. https://doi.org/10.1073/pnas.1910114117

13. Xu R, Yu P, Abramson MJ, Johnston FH, Samet JM, Bell ML, et al. Wildfires, global climate change, and human health. N Engl J Med. 2020;383(22):2173–2181. https://doi.org/10.1056/NEJMsr2028985

14. Gore A. Earth in the balance. New York: Houghton Mifflin; 1992.

15. Merle R. A guide to the financial crisis – 10 years later. The Washington Post. 10 September 2018. Available from: https://www.washingtonpost.com/business/economy/a-guide-to-the-financial-crisis--10-years-later/2018/09/10/114b76ba-af10-11e8-a20b-5f4f84429666_story.html

16. Witnessing a crime against humanity. The Guardian. 28 April 2021. Available from: https://www.theguardian.com/news/2021/apr/28/crime-against-humanity-arundhati-roy-india-covid-catastrophe?CMP=oth_b-aplnews_d-1

17. Harvey D. A brief history of neoliberalism. Oxford: Oxford University Press; 2005. https://doi.org/10.1093/oso/9780199283262.003.0010

18. MacMillan M. Which past is prologue? Heeding the right warnings from history. Foreign Affairs. Sept/Oct 2020. Available from: https://www.foreignaffairs.com/articles/united-states/2020-08-11/history-which-past-prologue

19. Rogers P. The crucial century, 1945-2045: Transforming food systems in a global context [Food Systems Academy – Transcript]. Available from: http://www.foodsystemsacademy.org.uk/videos/docs/Paul-Rogers-Transcript.pdf

20. Gray J. False dawn: The delusions of global capitalism. London: Granta; 1998.

21. Gill S, Bakker I, Wamsley D. Morbid symptoms, organic crises and enclosures of the commons. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 242–255 https://doi.org/10.1017/9781108692137.019

22. Homer-Dixon T. The American polity is cracked, and might collapse. Canada must prepare. The Globe and Mail. 31 December 2021 [updated 2022 Jan 02]. Available from: https://www.theglobeandmail.com/opinion/article-the-american-polity-is-cracked-and-might-collapse-canada-must-prepare/

23. Robinson WI. Global capitalism and the crisis of humanity. New York: Cambridge University Press; 2014. https://doi.org/10.1017/CBO9781107590250

24. Pearce F. As climate change worsens, a cascade of tipping points looms. YaleEnvironment360. 05 December 2019. Available from : https://e360.yale.edu/features/as-climate-changes-worsens-a-cascade-of-tipping-points-looms

25. Lenton TM, Held M, Kriegler E, Hall W, Lucht W, Rahmstorf S, et al. Tipping elements in the Earth's climate system. Proc Natl Acad Sci USA. 2008;105(6):1786–1793. https://doi.org/10.1073/pnas.0705414105

26. Benatar SR. Global justice and health: Re-examining our values. Bioethics. 2013;27(6):297–304. https://doi.org/10.1111/bioe.12033

27. Benatar S. A divided world in entropy. Society. 2018;55:200–206. https://doi.org/10.1007/s12115-018-0233-2

28. Greenblatt S. The swerve: How the world became modern. London: W. W. Norton & Co Ltd; 2001.

29. Baudot J, editor. Building a world community: Globalisation and the common good. Copenhagen: Royal Danish Ministry of Foreign Affairs; 2000.

30. Benatar SR, Upshur R. Virtue and values in medicine revisited: Individual health & global health. Clinical Med (Lond). 2014;14(5):495–499. https://doi.org/10.7861/clinmedicine.14-5-495

31. Benatar SR, Ashcroft R. International perspectives on resource allocation. In: Quah SR, Cockerham WC, editors. The international encyclopedia of public health. 2nd ed. Oxford: Academic Press; 2017. p. 316–321. https://doi.org/10.1016/B978-0-12-803678-5.00380-5

32. Ham C, Silva D. Integrating care and transforming community services: What works? Where next? [document on the Internet]. c2009 [cited 2022 Jan 30]. Available from: https://www.birmingham.ac.uk/Documents/college-social-sciences/social-policy/HSMC/publications/PolicyPapers/Policy-paper-5.pdf

33. Fairall LR, Bateman E, Cornick R, Faris G, Timmerman V, Folb N, et al. Innovating to improve primary care in less developed countries: Towards a global model. BMJ Innov. 2015;1:196–203. https://doi.org/10.1136/bmjinnov-2015-000045

34. McKee M. Health systems and health and healthcare reform. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 86–98. https://doi.org/10.1017/9781108692137.006

35. MccGwire M. The paradigm that lost its way. Int Affairs. 2001;77(4):777–803 https://doi.org/10.1111/1468-2346.00219

36. MccGwire M. Shifting the paradigm. Int Affairs. 2001;77(1):1–28. https://doi.org/10.1111/1468-2346.00236

37. Söder R, Mobjörk M, Lövebrand E. The Anthropocene and global politics: Rewriting the Earth as political space. Stockholm International Peace Research Institute WritePeace blog. 2020 September 02. Available from: https://www.sipri.org/commentary/blog/2020/anthropocene-and-global-politics-rewriting-earth-political-space

38. Westley F, Olsson P, Folke C, Homer-Dixon T, Vredenburg H, Loorbach D, et al. Tipping toward sustainability: Emerging pathways of transformation. Ambio. 2011;40(7):762–780. https://doi.org/10.1007/s13280-011-0186-9

39. Gill SR, Benatar SR. Reflections on the political economy of planetary health. Rev Int Political Econ. 2020;27(1):167–190. https://doi.org/10.1080/09692290.2019.1607769

40. Ottersen OP, Dasgupta J, Blouin C, Buss P, Chongsuvivatwong V, Frenk J, et al. The Lancet–University of Oslo commission on global governance for health. The political origins of health inequity: Prospects for change. Lancet. 2014;383(9917):630–667. https://doi.org/10.1016/S0140-6736(13)62407-1

41. Buchanan A, DeCamp M. Responsibility for global health. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 136–145. https://doi.org/10.1017/9781108692137.010

42. Gill S, Benatar SR. Global health governance and global power: A critical commentary on The Lancet University of Oslo Commission Report. Int J Health Serv. 2016;46(2):346–365. https://doi.org/10.1177/0020731416631734

43. Benatar SR, Lister G, Thacker SC. Values in global health governance. Glob Public Health. 2010;5(2):1–11. https://doi.org/10.1080/17441690903419009

44. Di Rugierro E. Global health governance for developing sustainability. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 440–449. https://doi.org/10.1017/9781108692137.037

45. Benatar S, Upshur R, Gill S. Understanding the relationship between ethics, neoliberalism and power as a step towards improving the health of people and our planet. Anthr Rev. 2018;5(2):155–176. https://doi.org/10.1177/2053019618760934

46. Galbraith JK. The affluent society. London: Penguin; 1958.

47. Galbraith JK. The good society: The humane agenda. London: Sinclair Stevenson; 1996.

48. Heilbroner R. Vision of the future: Distant past, yesterday, today, tomorrow. New York: Oxford University Press; 1995. https://doi.org/10.1093/acprof:oso/9780195102864.003.0002

49. Heilbroner R, Milberg W. The crisis of vision in modern economic thought. New York: Cambridge University Press; 1995. https://doi.org/10.1017/CBO9780511605574

50. Oreskes N, Conway EK. The collapse of Western civilization: A view from the future. New York: Columbia University Press; 2014. https://doi.org/10.7312/columbia/9780231169547.001.0001

51. Slaughter A-M, Laforge G. Opening up the order: A more inclusive international system. Foreign Affairs. March/April 2021.

52. Chagla Z, Pai M. COVID-19 boosters in rich nations will delay vaccines for all. Nature Med. 2021;27:1659–1660. https://doi.org/10.1038/s41591-021-01494-4

53. Mullard A. How COVID vaccines are being divvied up around the world. Nature News. 30 November 2020. https://doi.org/10.1038/d41586-020-03370-6

54. World Health Organization (WHO). Commission on the social determinants of health – Final report. Geneva: WHO; 2008. http://www.who.int/social_determinants/thecommission/finalreport/en/

55. Bakker I, Gill S, Wamsley D. Towards a new common sense: The need for new paradigms for global health beyond the COVID-19 emergency. In: Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 470–477. https://doi.org/10.1017/9781108692137.040

56. Del Ponte A, DeScioli P. Pay your debts: Moral dilemmas of international debt. Polit Behav. 2021. https://doi.org/10.1007/s11109-020-09675-6

57. Ganesan A, Vines A. Engine of war: Resources, greed, and the predatory state [document on the Internet]. c2004 [cited 2022 Jan 30]. Available from: http://www.hrw.org/legacy/wr2k4/14.htm

58. Hinton J, Maclurcan D. A not-for-profit world beyond capitalism and economic growth. Ephemera: Theory & Politics in Organization. 2017;17(1):147–166.

59. The Blue Dot 1: Exploring new ideas for a shared future. New Delhi: UNESCO MGIEP; 2015. https://mgiep.unesco.org/the-blue-dot-issue-1

60. Brecher J, Costello T, Smith B. Globalization from below: The power of solidarity. Cambridge, MA: South End Press; 2000.

61. Albrecht G. Exiting the Anthropocene and entering the Symbiocene. Minding Nature. 2016;9(2):12–16. Available from: https://www.humansandnature.org/filebin/pdf/minding_nature/may_2016/Albrecht_May2016.pdf

62. Castles S, Davidson A. Citizenship and migration: Globalisation and the politics of belonging. London: MacMillan Press Ltd; 2000.

63. Potter VR. Bioethics: Bridge to the future. Hoboken, NJ: Prentice-Hall Inc.; 1971.

64. Kung H. Global responsibility: In search of a new world ethic. New York: Continuum; 1995.

65. Benatar S, Daibes I, Tomsons S. Inter-philosophies dialogue: Creating a paradigm for global health ethics. Kennedy Inst Ethics J. 2016;26(3):323–346. https://doi.org/10.1353/ken.2016.0027

66. Bloomberg Editorial Board. The U.S. doesn't need new missiles in the ground. Bloomberg. 22 February 2021. Available from: https://www.bloomberg.com/opinion/articles/2021-02-22/the-u-s-doesn-t-need-new-missiles-in-the-ground

67. Alstrope T, Fishel S. Horror, apocalypse and world politics. Int Affairs. 2020;96(3):631–648. https://doi.org/10.1093/ia/iiaa008

68. Garrett L. Betrayal of trust: The collapse of public health. New York: Hyperion; 2000.

69. Zywert K, Quilley S, editors. Health in the Anthropocene: Living well on a finite planet. Toronto: Toronto University Press; 2020. https://doi.org/10.3138/9781487533410

70. Rischard JF. High noon: 20 Global problems, 20 years to solve them. New York: Basic Books; 2020.

71. Milman O. Global dismay as supreme court ruling leaves Biden's climate policy in tatters. The Guardian. 06 July 2022. Available from: https://www.theguardian.com/environment/2022/jul/06/supreme-court-epa-ruling-biden-climate-policy-global-reaction

72. Barbour IG, editor. Western man & environmental ethics. Reading, MA: Addison-Wesley Pub. Co.; 1973.

73. Oreskes N, Conway EM. Merchants of doubt. New York: Bloomsbury Press; 2010.

74. Speth JG. They knew: The US Federal Government's fifty-year role in causing the climate crisis. Cambridge, MA: MIT Press; 2021. https://doi.org/10.7551/mitpress/14083.001.0001

75. U.S. Energy Information Administration. Monthly Energy Review. October 2020. DOE/EIA-0035(2020/10). Available from: https://www.eia.gov/totalenergy/data/monthly/pdf/mer.pdf

76. Carbon Tracker. The sky's the limit: Solar and wind energy potential is 100 times as much as global energy demand. London/New York: Carbon Tracker; 2021. Available from: https://carbontracker.org/reports/the-skys-the-limit-solar-wind

77. Gill S, editor. Global crisis and the crisis of global leadership. Cambridge, UK: Cambridge University Press; 2012. https://doi.org/10.1017/CBO9781139046596

78. Mills CW. The sociological imagination (40th Anniversary ed.). New York: Oxford University Press; 1999.

79. Glover J. Humanity: A moral history of the twentieth century. New Haven, CT: Yale Nota Bene; 2001.

80. Fukuyama F. The pandemic and political order. It takes a State. Foreign Affairs. July/August 2000. Available from: https://www.foreignaffairs.com/print/node/1126047

81. "The American way of life is not up for negotiation. Period." — George H. W. Bush. QuotePark. Available from: https://quotepark.com/quotes/2033297-george-h-w-bush-the-american-way-of-life-is-not-up-for-negotiation/

82. Smil V. How the world really works. The science behind how we got here and where we are going. London: Viking; 2022.

83. Kahneman D. Sixty four inspirational Daniel Kahneman quotes on success in life. Available from: https://www.overallmotivation.com/quotes/daniel-kahneman-quotes/

84. Benatar SR. Annual Human Rights Lecture University of Alberta. Global health & human rights: Working with the 20th century legacy. Available from: https://www.ualberta.ca/global-education/visiting-lectureship-human-rights/past-lecturers

85. Perkins J. Hoodwinked. New York; Broadway Books; 2009.

86. The 1619 project. New York Times Magazine. 20 April 2000.

87. Yearby R. Structural racism and health disparities: Reconfiguring the social determinants of health framework to include the root cause. J Law Med Ethics. 2020;48:518–526. https://doi.org/10.1177/1073110520958876

88. Bailey ZD, Feldman JM, Bassett MT. How structural racism works - racist policies as a root cause of U.S. racial health inequities. N Engl J Med. 2021;384:768–773. https://doi.org/10.1056/NEJMms2025396

89. Hosle V. The third world as a philosophical problem. Social Research. 1992;52:227–262.

90. Wallerstein I. Globalization or the Age of Transition?: A long-term view of the trajectory of the world-system. Int Sociol. 2000;15(2):249–265. https://doi.org/10.1177/0268580900015002007

91. Van Rensburg HCJ, Benatar S. The legacy of apartheid in health & health care. S Afr J Sociol. 1993;24(4):99–111. https://doi.org/10.1080/02580144.1993.10431680

92. Tomsons S, Mayer L. Philosophy and Aboriginal rights. Oxford: Oxford University Press; 2013.

93. Benatar S, Brock G, editors. Global health: Ethical challenges. 2nd ed. Cambridge, UK: Cambridge University Press; 2021. p. 326–369. https://doi.org/10.1017/9781108692137

94. Bierman F. The future of 'environmental' policy in the Anthropocene: Time for a paradigm shift. Environ Politics. 2021;30(1–2):61–68. https://doi.org/10.1080/09644016.2020.1846958

95. Bakker I, Gill S, editors. Power, production & social reproduction. Toronto: Palgrave MacMilllan; 2003. https://doi.org/10.1057/9780230522404

96. Galtung J. The true worlds. New York: The Free Press; 1980.

97. Packer G. Making America again. The Atlantic. 11 September 2020. Available from: https://www.theatlantic.com/projects/making-america-again/

98. Sachs JD. Blood in the sand. Project Syndicate. 17 August 2021. Available from: https://www.project-syndicate.org/commentary/afghanistan-latest-debacle-of-us-foreign-policy-by-jeffrey-d-sachs-2021-08?utm_source=Project+Syndicate+Newsletter&utm_campaign=8aaa27b08b-sunday_newsletter_22_08_2021&utm_medium=email&utm_term=0_73bad5b7d8-8aaa27b08b-104754877&mc_cid=8aaa27b08b&mc_eid=40cbc00f41

99. Sachs JD. America's confrontational foreign policy failed. It should pursue a cooperative global policy. Boston Globe. 3 September 2021. Available from: https://www.bostonglobe.com/2021/09/03/opinion/americas-confrontational-foreign-policy-failed-it-should-pursue-cooperative-global-policy/

100. JM Keynes quotes [webpage on the Internet]. No date [cited 2022 Jan 30]. Available from: https://www.brainyquote.com/quotes/john_maynard_keynes_385471

101. Hegel GWF. Lectures on the philosophy of history [webpage on the Internet]. No date [cited 2022 Jan 30]. Available from: https://quotepark.com/works/lectures-on-the-philosophy-of-history-7557/

102. Whyte D, Wiegratz J. Neoliberalism and the moral economy of fraud. New York: Routledge; 2017. https://doi.org/10.4324/9781315680545

103. McCarthy J. Fossil fuel subsidies exceed $5 trillion annually. Here's why. Global Citizen. 08 December 2017. Available from: https://www.globalcitizen.org/en/content/fossil-fuel-subsidies-trillions-climate-change/?gclid=EAIaIQobChMIlc3txv699AIVEQGLCh1TgwTIEAAYAyAAEgII

104. Bednarz D, Beavis A. Neoliberalism, degrowth and the fate of health systems. Resilience. 14 September 2012. Available from: https://www.resilience.org/stories/2012-09-14/neoliberalism-degrowth-and-fate-health-systems/

105. Kapuściński R. The other. London: Verso; 2018.

106. Jansen J. As by fire: The end of the South African university. Cape Town: Tafelberg; 2017.

107. Tomaselli KG. Contemporary campus life. Cape Town: Best Red; 2021.

108. Ansloos JP. The medicine of peace: Indigenous youth decolonizing healing and resisting violence. Halifax & Winnipeg: Fairwood Publishing; 2017.

109. Diamond J. Upheaval: How nations cope with crisis and change. London: Penguin; 2019.

110. Gates B. How to avoid a climate disaster. New York: Random House; 2021.

111. Rees M. On the future prospects for humanity. Princeton, NJ: Princeton University Press; 2018.

112. Ferguson N. Doom: The politics of catastrophe. New York: Penguin; 2021.

113. Graeber D, Wengrow D. The dawn of everything: A new history of humanity. Toronto: Signal; 2021.

114. Kolbert E. Under a white sky: The nature of the future. New York; Crown; 2021.

115. Mbeki M. Architects of poverty: Why African capitalism needs changing. Johannesburg: Picador; 2009.

116. Johnson RW. Fighting for the dream. Cape Town: Jonathan Ball; 2019.

117. Judicial Commission of Inquiry into Allegations of State Capture, Corruption and Fraud in the Public Sector Including Organs of State Report: Part VI Vol. 4: Summary of Recommendations, Chairperson: Justice R.M.M Zondo Chief Justice of the Republic of South Africa. Available from: https://www.gcis.gov.za/sites/default/files/[Electronic]%20State%20Capture%20Commission%20Report%20Part%20VI%20Vol%20IV.pdf

118. Emmott B. The fate of the West: The battle to save the world's most successful political idea. London: Profile Books Ltd; 2017.

119. Lakoff G, Johnson M. The metaphors we live by. Chicago, IL: University of Chicago Press; 1980.

120. Smart N. World views: Cross-cultural explorations of human beliefs. Princeton, NJ: Prentice-Hall; 1995.

Note: A list of additional readings is provided in the Supplementary material.

AUTHORS:
Dirk Brand[1]
Jerome A. Singh[2]
Annelize G. Nienaber McKay[3]
Nezerith Cengiz[4]
Keymanthri Moodley[4]

AFFILIATIONS:
[1]School of Public Leadership, Stellenbosch University, Stellenbosch, South Africa
[2]School of Law, University of KwaZulu-Natal, Pietermaritzburg, South Africa
[3]Division of Law, Abertay University, Dundee, Scotland, United Kingdom
[4]Centre for Medical Ethics and Law, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Nezerith Cengiz

EMAIL:
ncengiz@sun.ac.za

DATES:
Received: 06 May 2022
Revised: 24 Aug. 2022
Accepted: 19 Oct. 2022
Published: 26 Oct. 2022

HOW TO CITE:
Brand D, Singh JA, Nienaber McKay AG, Cengiz N, Moodley K. Data sharing governance in sub-Saharan Africa during public health emergencies: Gaps and guidance. S Afr J Sci. 2022;118(11/12), Art. #13892. https://doi.org/10.17159/sajs.2022/13892

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITOR:
Pascal Bessong

KEYWORDS:
data transfer regulations, data sharing, public health, sub-Saharan Africa

FUNDING:
US National Institutes of Health (1U01MH127704-01)

Data sharing governance in sub-Saharan Africa during public health emergencies: Gaps and guidance

While the COVID-19 pandemic has captured the attention of the global community since the end of 2019, deadly health pandemics are not new to Africa. Tuberculosis (TB), malaria and human immunodeficiency virus (HIV) count amongst other serious diseases that have had a catastrophic impact on the African continent. Effective responses to such pandemics require high-quality, comprehensive data sets that can inform policymaking and enhance healthcare decision-making. While data is driving the information economy in the 21st century, the scarcity in Africa of carefully curated, large epidemiologic data sources and analytical capacity to rapidly identify and understand emerging infectious diseases poses a major challenge to mounting a time-sensitive response to unfolding pandemics. Data access, sharing and transfer between countries are crucial to effectively managing current and future health pandemics. Data access and sharing, however, raises questions about personal privacy, the adequacy of governance mechanisms to regulate cross-border data flows, and ethical issues relating to the collection and use of personal data in the interests of public health. Sub-Saharan Africa's most research-intensive countries are characterised by diverse data management and privacy governance frameworks. Such regional variance can impede time-sensitive data sharing and highlights the need for urgent governance reforms to facilitate effective decision-making in response to rapidly evolving public health threats.

Significance:

We explore governance considerations that ought to apply to the collection, transfer, and use of data in public health emergencies. Specifically, we provide an overview of the prevailing data sharing governance landscape in selected African countries. In doing so, we identify limitations and gaps that impede effective data collation, sharing and analysis. This work could find utility amongst a range of stakeholders, including bioinformaticians, epidemiologists, artificial intelligence coders, and government decision-makers. While this work focuses primarily on an African context, the issues explored are of universal concern and therefore of relevance to a broader international audience.

Introduction

The collation, storage, and sharing of personal data are becoming increasingly important in public health research and surveillance. Balancing data protection and privacy concerns with data sharing and open science is a persistent challenge. The COVID-19 pandemic has highlighted the importance of data sharing in the public interest.[1,2]

While the world has afforded significant attention to the COVID-19 pandemic since the end of 2019, Africa has historically been the perennial epicentre of some of the world's deadliest diseases, including tuberculosis, malaria, and human immunodeficiency virus (HIV). The clinical and public health management of these pandemics requires the collation, storage, and use of personal and aggregated health data to inform clinical decision-making, scientific endeavours, and policymaking. The COVID-19 pandemic has demonstrated the need for transparency – not just to keep the public informed about the nature and spread of the disease, but also to enable citizens to hold governments accountable for the extraordinary measures they have adopted to manage the pandemic. Seen in this light, data sharing is also in the public interest as the guidance and directives the public is expected to follow, should be evidence-based.

The rapid global spread of COVID-19 has necessitated greater international cooperation. Although optimised data collection at a national level is important to strengthen the domestic response to COVID-19, regional and international cooperation in accessing and using such data to inform policymaking beyond the data source setting is equally important due to the global impact of the disease.[3,4] Not surprisingly, the World Health Organization (WHO) has repeatedly stressed the need for cooperation between countries to end the COVID-19 pandemic.[5]

At a very basic level, there should be appropriate infrastructure in a country for the effective collection of personal information of data subjects, and the storage thereof. Moreover, a country should possess analytical capacity to understand the implications of the data. Limited Internet access, poor quality data sources, and inadequate analytical capacity characterise many African settings.[6] In the context of COVID-19, such shortcomings contributed to critical knowledge gaps, such as the impact of COVID-19 on African children.[7] The scale of the current COVID-19 pandemic has underscored the need for appropriate data collation, storage, and analysis globally to strengthen evidence-based approaches to managing the pandemic.[6,8] Although many African countries have promulgated data protection legislation, the practical gap between the availability of data sources and analytical capacity is of serious concern. While such gaps negatively impact local and national efforts to fight the pandemic, they also underscore the global inequity regarding the resources and ability to manage health pandemics.[9,10]

Data requirements to manage COVID-19 and other health pandemics

The lack of harmonised data sets and information systems poses a major barrier to the effective management of public health emergencies. Moreover, time-sensitive access to data may not always be possible.[11,12] Access

to large, high-quality data sets is also crucial to the development of artificial intelligence (AI) models, which can help optimise predictive decision-making in health pandemics.[13] One of the basic building blocks in managing health pandemics at a country level is the development of national health information systems that include both public and private sources of health data.

Broad categories of data are required to effectively respond to pandemics. Primary data sources include, but are not limited to, mobility surveillance data, incidence, hospitalisation, recovery, and mortality data aggregated by amongst other factors, age, sex, comorbidity, and vaccination status. Epidemiological data and omics research data are, from a research perspective, critical data sources that support an evidence-based approach to managing a pandemic.[14] An inability to gain timely access to data, trust deficits, diverse data sources and diverse data needs are amongst the factors that can hinder an effective response to a pandemic (Box 1).[15-18] To overcome such barriers, the characterisation and formats of various health data sources should be harmonised and standardised to facilitate sharing amongst relevant stakeholders, such as researchers, public health officials, and international agencies.

Box 1: Key limitations in data sharing

- Timely access to data during a public health emergency is critical to gain detailed knowledge of a pandemic that could assist in developing effective public health responses. Many low- and middle-income countries have limited capacity to undertake epidemiological, clinical, and other research, thus negatively impacting on the ability to respond effectively to public health emergencies.
- Data sharing, including cross-border transfer of data, is limited due to a lack of trust about the future use of the data, insufficient data transfer regimes or the academic competition to be the first to publish results of scientific research and thus unwillingness to share research data.
- Data needs of different stakeholders differ, which makes a common approach to data sharing difficult.[17]
- There is a variety of types of data collected and used in the context of public health, which complicates effective data sharing practices.
- In cases where data quality is questionable or the data is incomplete, the scope of using the data in the management of a health pandemic is limited.
- Colonial legacies and social and economic inequalities could have a negative impact on international cooperation to share data due to questions about trust between international partners.[18]
- A lack of standardisation of types of data limits international cooperation.[15]

In considering data sharing needs during a pandemic, it is important to reflect on how 'open data' could facilitate access to, and use of, health data. The scale of the COVID-19 pandemic and the need to have time-sensitive epidemiological data publicly accessible on a daily basis has boosted support for open data. An important motivating factor for open data is the opportunity such data creates for a diverse group of stakeholders, including researchers and software developers, to analyse the data and generate new insights and applications.[19] The internationally renowned COVID-19 global dashboard created by the Johns Hopkins University[20] and other similar country-level dashboards that are updated daily, count as examples of how open data can add value to the management of pandemics. Open data, however, also present challenges and potential risks. Some commentators note that government efforts to provide and maintain open data are costly and require considerable expertise, which is not universally available.[19]

Data quality, access to primary data and completeness of data are additional challenges to open data. Data access governance, intellectual property rights and privacy considerations also merit noting.[21] When open data applications are created, data anonymisation is usually part of that creation process. However, privacy concerns could arise if there remains a possibility to de-anonymise personal data, notwithstanding technological tools, such as differential privacy used in machine learning modelling, which could be used to respond to such privacy concerns. Despite such potential risks and challenges, the benefit of timely open data during a pandemic is clear. The COVID-19 pandemic has demonstrated the key role that bioinformaticians and epidemiologists can play in collating and analysing data sets to rapidly inform pandemic decision-making.[22] Such contributions would not be possible without the collection of, and time-sensitive access to, large, high-quality data sets.

Initiatives to facilitate cross-border data sharing in Africa

In 2002, the WHO published an Integrated Disease Surveillance and Response (IDSR) for WHO's African region, which was widely adopted by African member states.[23] The African Union (AU) is similarly leading an effort to harmonise statistical data sets to facilitate and foster data sharing in Africa.[24] In 2018, the Africa Centres for Disease Control (Africa CDC), which serves as a specialised technical institution of the AU, published a Framework for Event-based Surveillance, which is intended to complement and enhance the implementation of IDSR.[25] In publishing the third edition of its IDSR Guidelines in 2019, the WHO conceded that progress towards a coordinated, integrated surveillance system in Africa has been mixed.[23] In recognition thereof, the Africa CDC has stepped up its efforts to support Member States to develop and establish high-quality public health information and technology systems.

One of the Africa CDC's flagship initiatives involves the development of a continental wide public health information system platform by linking public health institutes in each country through a wide area network managed by the Africa CDC.[26] The strategic objective of this initiative includes, amongst others, enhancing secure electronic transmission of relevant data and reports, facilitating the development and promotion of network domains, and adopting informatics guidelines and standards to enable interconnectivity and electronic transmission of data and information among Africa CDC institutes.[26]

Similarly, the Africa CDC has also established the Regional Integrated Surveillance and Laboratory Network (RISLNET) initiative, which aims to establish an integrated electronic network of regional surveillance platforms by leveraging existing regional public health assets, such as the surveillance and laboratory networks operated by public agencies, private organisations, foundations, and universities in eastern and southern Africa.[27] Data from RISLNET is intended to inform, amongst others, the Extension for Community Healthcare Outcomes (ECHO) platform, which aims to share critical, timely, lifesaving information and data with healthcare workers at different geographical locations.[28] Digital disease surveillance is described as 'the aggregation and analysis of data available on the internet, such as search engines, social media, and mobile phones, and not directly associated with patient illnesses or medical encounters'[29]. In recognition of its potential to contribute to disease surveillance, the Africa CDC has established a pilot digital disease surveillance programme aimed at conducting real-time surveillance of infectious diseases in Africa by monitoring social media and building capacity in 'Big Data' approaches for outbreak prediction, analysis, and prevention.[29] Such ambitious regional initiatives speak to the need for time-sensitive cross-border data sharing but also raise privacy concerns.

Data stewardship: Ethics and governance considerations

The Open Data Institute characterises data sharing in three aspects: (1) stewardship of data (collection, maintenance, sharing); (2) creating information from the data (analysis, insights); and (3) making informed decisions utilising data from different sources.[30] Various technical, motivational, economic, political, legal, and ethical barriers can negatively impact on public health data sharing initiatives.[31] Such barriers include a

lack of resources in the public sector (economic barrier), ownership and copyright (legal barrier) and lack of reciprocity in data sharing practices (ethical barrier). The COVID-19 pandemic has underscored the need to develop a universal system or standard for collating and sharing data, including research outputs based on such data.[32-34] This universality is especially crucial during the initial stages of an emerging pandemic when information about the nature and spread of a disease should be shared internationally as soon as possible, as well as later, when coordinated evidence-based efforts to fight the pandemic are important. In 2022, the OECD published recommendations on enhancing access to and sharing of data[12], which include the principles and guidelines captured in Box 2. In a research report[24] that focuses on data sharing to enhance public health in Africa, equitable, ethical, and efficient data sharing constitute key principles.

Box 2: Recommendations of the Council on Enhancing Access to and Sharing of Data in 2022

- Promote trustworthiness of the data ecosystem.
- Enhance transparency of data access and sharing arrangements.
- Incentivise data access and sharing.
- Foster effective and responsible data access, sharing and use across society.

The United Kingdom (UK) government has followed a practical approach in devising its Data Ethics Framework (DEF), which offers guidance for responsible data use in the public sector.[35] The DEF posits three principles, namely transparency, accountability, and fairness. In this context, transparency means that data processes are open to inspection and that information about a project must be published in an understandable and accessible format. Adherence to accountability means that governance and oversight mechanisms are in place and implemented effectively. To mitigate any bias or discrimination in the capturing, sharing and use of data, the principle of fairness must apply, that is, there must be respect for the dignity of individuals and an aim to deliver fair and non-discriminatory outcomes. Despite being UK-focused, DEF's guidance principles are universally relevant and should be considered in the context of cross-border data transfers between African countries. Transparency, accountability, and fairness are also legal principles underpinning administrative and constitutional law. Although they are not necessarily included in data protection legislation, they are still important in the context of responsible data use and should guide the drafting of data transfer agreements.

Some commentators have argued for a wide approach to drafting a data transfer agreement in a research context, suggesting that it should include, amongst others, provisions on ethical considerations including ethical approval, and benefit sharing.[36] Novel ideas such as benefit sharing could be considered from a research ethics perspective, but the *Protection of Personal Information Act* (POPIA) and other data protection legislation are aimed at protecting personal information and not aimed at research ethics per se. Although the detail of a data transfer agreement is context specific, and anchored in the relevant data protection legislation, any data transfer agreement in a research context should, at minimum, include the elements indicated in Box 3. In some settings, national data protection authorities may need to prospectively approve cross-border transfers involving personal data.

The Research Data Alliance (RDA) COVID-19 Working Group has published a set of recommendations and guidelines relating to the collection and sharing of data in the context of COVID-19.[11,12] The work of the RDA COVID-19 Working Group is divided into four categories, namely clinical research, omics, epidemiology, and social sciences, all of which contribute to the multidisciplinary nature of managing a pandemic. In all these areas there is a need for more data sharing, but often a lack of proper data sharing or data transfer agreements negatively impacts the analytical work as well as the policy responses as part of the management of a pandemic.

Box 3: Minimum requirements for a data transfer agreement

- Responsibilities of the provider and recipient of the data
- Purpose of the use of the data
- Description of the data
- Time period of the agreement
- Access to the data (e.g. in a research context there could be various people in the working environment of the recipient who need access to the data)
- Confirmation of adherence to all the legal requirements for lawful processing of data
- Publication of the research results based on the transferred data
- Ethical clearance for the use of the data in the research
- Dispute resolution provisions

Africa's data sharing regulatory landscape

Data protection laws provide the legal framework for the collection, access to and sharing of data, as well as the cross-border transfer of data. Many African countries have embarked on initiatives to regulate data protection. In the fields of public health, environmental and occupational health, South Africa, Nigeria, Kenya, Ethiopia and Uganda[37] rank respectively as the most research-intense countries in sub-Saharan Africa by research output[32-34]. It is thus apt to briefly consider how these settings manage cross-border data transfers.

Table 1 categorises the rigour of national data protection laws regarding cross-border transfer of personal data.[38] Table 1 is not aimed at providing a strict overall categorisation of various data protection laws, but rather, is focused on the scope of legal protection afforded to data subjects in relation to the cross-border transfer of their personal data. Countries with *stringent* rules require notification of, or approval by, a relevant data protection authority, and/or special conditions (such as proof of appropriate safeguards with respect to the protection and security of personal data), as well as consent from the data subject.

South Africa and Kenya count amongst countries that could be described as providing stringent data export protection to data subjects. For example, *Kenya's Data Protection Act of 2019* complies with the European Union (EU) legal standards, which are generally regarded as being stringent in nature. For data to be transferred out of Kenya, the data processor must verify to the data commissioner that the third-party recipient's jurisdiction is bound by appropriate safeguards for the security and protection of the data. It is also important that the data transfer be purposeful, such as being necessary for the conclusion or performance of a contract or legal claim, and the public or data subjects' interests. In addition, consent from the data subject is also required for cross-border data transfers.[38]

Countries falling in the *moderate* category allow for more than one possible legal ground to permit data export, such as consent of the data subject, but do not require notification or approval by the data protection authority. Nigeria counts amongst countries providing moderate data export protection to data subjects as the country's data protection law does not require third-party recipients of data to be bound by adequate data protection law, agreements, or corporate rules if the data subject provides consent after being informed of possible risks of inadequate data protection or if the transfer meets a certain exception. One example of such exception is the public's or data subject's interest. Beyond obtaining consent from data subjects for data transfers, the Nigeria Data Protection Regulation 2019 requires the National Information Technology Development Agency (NITDA) or Honourable Attorney General of the Federation (HAGF) to ensure that the third-party recipients of the transferred data have adequate data protection standards in place.[38]

Ghana's data protection legislation does not contain any provisions pertaining to cross-border transfer of personal information[39] and could

Table 1: Data protection laws, regarding cross-border transfer of personal data, of sub-Saharan African countries ranked by research output in "Public Health, Environment, and Occupational Health"[37]

Rank and country	Legal requirements	Legislation	Data export protection classification
South Africa	A responsible party may only transfer personal data outside South Africa if the recipient is subject to a law, binding corporate rules or binding agreement that provides adequate protection; or the data subject consents to the transfer; or the transfer is necessary in terms of the provisions of the Act.	Section 72 of the *Protection of Personal Information Act, 4 of 2013*	Strict
Nigeria	Cross-border transfer of personal data is subject to authorisation by the Attorney General or the National Information Technology Development Agency (NITDA) based on an adequate level of protection. In the absence of authorisation by the Attorney General or the NITDA, personal data transfer may only take place if the data subject gives consent, or the data transfer is necessary in terms of the Regulation.	Reg. 2.11 and 2.12 of the Nigeria Data Protection Regulation, 2019	Moderate
Kenya	Data transfer is allowed only if there is proof of adequate data protection safeguards or consent from the data subject. The data controller or data processor must provide proof of appropriate safeguards to the Data Commissioner. The data transfer must be necessary in terms of the Act.	Section 25(h) 48 of the *Data Protection Act, No. 24 of 2019* (Kenya)	Strict
Ethiopia	Cross-border data transfer may only take place subject to an adequate level of data protection in the recipient country. The data controller or data processor must provide proof to the Data Protection Commission of an appropriate level of protection, or the data subject must give consent to the proposed transfer, or the transfer must be necessary, or the transfer must be made from a register and intended to provide information to the public.	Sections 27–30 of the Draft Proclamation to Provide for Personal Data Protection, 2021 (Ethiopia)	Strict
Uganda	The data processor or data controller must ensure that there are adequate measures in place for the protection of personal data, or the data subject must provide consent.	Section 19 of the *Data Protection and Privacy Act, 2019* (Uganda)	Strict

thus be described as providing *inadequate* protection to data subjects in relation to the export of their personal data.

The diverse legal landscape governing data sharing in sub-Saharan Africa – including the stringency of data export provisions – highlights that cross-border data transfers will have to be evaluated on a case-by-case basis as there is no uniform law across the continent akin to the General Data Protection Regulation (GDPR 2018), which constitutes a common legal framework for all EU Member States. Although the AU Commission is developing a data policy framework for Africa to harness digital technologies and innovation in an attempt to bridge the digital divide, this process is ongoing and will take time to implement.[16] Further, the AU Convention on Cyber Security and Personal Data (2014) has been ratified by only 13 AU Member States.[40]

It is important to note that in addition to specific data protection legislation, 29 sub-Saharan countries have published some form of research ethics guidance regarding the collection and use of human biological specimens, which are essential in national health systems and important to consider in international research cooperation.[41] These countries are Benin, Botswana, Burkina Faso, Cameroon, Democratic Republic of the Congo, Equatorial Guinea, Ethiopia, Gabon, The Gambia, Ghana, Guinea, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritius, Mozambique, Nigeria, Rwanda, Senegal, Sierra Leone, South Africa, Sudan, Tanzania, Uganda, Zambia, and Zimbabwe.[41]

These guidance documents do not specifically govern data sharing. In South Africa, the Academy of Science of South Africa (ASSAf) – a statutory body and the country's apex science advisory body, has developed a privacy Code of Conduct for Research. The Code is intended to provide guidance to the research community about the use of data in research and is binding.[42] The Code reiterates the country's legislative stance on the transfers of personal information outside the country.

It is evident from the above that data sharing governance varies across Africa. Adequacy of legal protection for cross-border transfers is of particular importance when scientists and other stakeholders in different countries need to share data to effectively respond to a rapidly evolving pandemic. Cross-border data transfers of personal information may only take place if certain requirements are met in accordance with the applicable local legislation. A common approach found in various data protection laws includes the imposition of certain conditions, as indicated in Box 4.

Box 4: Common conditions enforced in various data protection laws

- The third-party recipient of the data is subject to a law, binding corporate rules or a binding agreement, which provides adequate protection.
- The data subject gives consent for the cross-border transfer.
- The transfer is necessary for the performance of a contract between the data subject and the responsible party/controller.
- The transfer is necessary for the conclusion or performance of a contract in the interest of the data subject between the responsible party and a third party.
- The transfer is for the benefit of the data subject.

Data transfer agreements can facilitate the cross-border transfer of data and catalyse international research collaboration.[17] A lack of pre-approved data sharing agreements and archaic health data systems count amongst the critical shortcomings of the global community's response to the COVID-19 pandemic.[17,26] To avoid similar shortcomings in the future, relevant authorities should develop standardised tools and templates for international research collaboration. The development of such tools and templates should be based on sound governance and ethics principles. Additional factors to consider in the processing of personal information are outlined in Box 5.

Box 5: Factors to consider in the processing of personal information[18]

- The need for a time limit for the retention of the data
- Clarity and a sound legal basis for the purpose of the data processing
- Proportionality of the measures taken in processing the personal data
- Transparency and explainability
- Accountability
- Integration of privacy by design
- Realisation of data protection impact assessments

Conclusion and recommendations

This article highlights the factors that impact data sharing in sub-Saharan Africa, especially in the context of managing health pandemics. The COVID-19 pandemic has underscored the need for a reliable and accessible data ecosystem that could inform the management of public health threats. The combined effect of diverse limitations or barriers to data sharing in public health necessitates more dedicated continental and international cooperation as well as the development of standard formats for health data. Harmonised data sources and their integration into national health information systems will create a comprehensive data set that includes epidemiologic, clinical as well as behavioural data relating to public health emergencies. Such a holistic approach to data management should underpin evidence-based decision-making. The principles of transparency, fairness and accountability should underpin the development of a reliable and accessible data ecosystem. To facilitate cross-border data transfers involving personal data, standard contractual provisions and templates for cross-border data transfers should be developed by data protection authorities in Africa. This will facilitate not just scientific cooperation between countries, but also an integrated cross-border approach to the management of future pandemics.

Acknowledgement

We acknowledge the US National Institutes of Health (NIH) DS-I Africa for funding (grant 1U01MH127704-01).

Competing interests

We have no competing interests to declare.

Authors' contributions

All authors made substantial contributions to the conception or design of the work or the acquisition, analysis, or interpretation of data for the work; drafted/revised the work critically for important intellectual content; approved the final version; and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. The World Health Organization. World health report 2013: Research for universal health coverage [webpage on the Internet]. c2013 [cited 2022 Aug 22]. Available from: https://www.afro.who.int/publications/world-health-report-2013-research-universal-health-coverage
2. The World Medical Association (WMA). WMA Statement on Healthcare Information for All [webpage on the Internet]. c2019 [cited 2022 Aug 22]. Available from: https://www.wma.net/policies-post/wma-statement-on-healthcare-information-for-all/
3. Do Lee W, Qian M, Schwanen T. The association between socioeconomic status and mobility reductions in the early stage of England's COVID-19 epidemic. Health Place. 2021;69, Art. #102563. https://doi.org/10.1016/j.healthplace.2021.102563
4. Trestian R, Celeste E, Xie G, Lohar P, Bendechache M, Brennan R, et al. The privacy paradox – investigating people's attitude towards privacy in a time of COVID-19. In: Proceedings of the 14th International Conference on Communications (COMM); 16–18 June 2022; Bucharest, Romania. IEEE; 2022. p. 1–6. https://doi.org/10.1109/COMM54429.2022.9817170
5. Commonwealth Secretariat, World Health Organization (WHO). Commonwealth and WHO to strengthen cooperation on health, including access to vaccines [media release]. World Health Organization (WHO). 7 February 2022. Available from: https://www.who.int/news/item/07-02-2022-commonwealth-and-who-to-strengthen-cooperation-on-health-including-access-to-vaccines
6. Nachega JB, Uthman OA, Ho YS, Lo M, Anude C, Kayembe P, et al. Current status and future prospects of epidemiology and public health training and research in the WHO African region. Int J Epidemiol. 2012;41(6):1829–1846. https://doi.org/10.1093/ije/dys189
7. Sam-Agudu NA, Rabie H, Pipo MT, Byamungu LN, Masekela R, Van der Zalm MM, et al. The critical need for pooled data on coronavirus disease 2019 in African children: An AFREhealth call for action through multicountry research collaboration. Clin Infect Dis. 2021;73(10):1913–1919. https://doi.org/10.1093/cid/ciab142
8. Haleem A, Javaid M, Khan IH, Vaishya R. Significant applications of big data in COVID-19 pandemic. Indian J Orthop. 2020;54(4):526–528. https://doi.org/10.1007/s43465-020-00129-z
9. Townsend B. The lawful sharing of health research data in South Africa and beyond. Inf Commun Technol Law. 2022;31(1):17–34. https://doi.org/10.1080/13600834.2021.1918905
10. Staunton C, Tschigg K, Sherman G. Data protection, data management, and data sharing: Stakeholder perspectives on the protection of personal health information in South Africa. PLoS ONE. 2021;16(12), e0260341. https://doi.org/10.1371/journal.pone.0260341
11. RDA COVID-19 Working Group. RDA COVID-19 Recommendations and guidelines on data sharing. Research Data Alliance (RDA); 2020. https://doi.org/10.15497/rda00052
12. Research Data Alliance. Enhancing access to research data during crises: Lessons learned from the COVID-19 pandemic. Paris: Organisation for Economic Co-operation and Development (OECD); 2021. Available from: https://www.oecd.org/sti/inno/enhance-access-research-data-during-crises.htm
13. Syrowatka A, Kuznetsova M, Alsubai A, Beckman AL, Bain PA, Craig KJT, et al. Leveraging artificial intelligence for pandemic preparedness and response: A scoping review to identify key use cases. NPJ Digit Med. 2021;4(1), Art. #96. https://doi.org/10.1038/s41746-021-00459-8
14. Tonacci A, Genovese S, Pioggia G, Gangemi S. COVID-19 pandemic: Different roles for scientific publications and funding face to epidemiological data – an European, country-based perspective. Clin Mol Allergy. 2021;19(1), Art. #16. https://doi.org/10.1186/s12948-021-00154-9
15. Wellcome Trust. Data sharing in public health emergencies [webpage on the Internet]. c2019 [cited 2022 Aug 22]. Available from: https://wellcome.org/what-we-do/our-work/data-sharing-public-health-emergencies
16. African Union (AU). The digital transformation strategy for Africa (2020-2030). Addis Ababa: AU; 2021. Available from: https://au.int/sites/default/files/documents/38507-doc-dts-english.pdf
17. The Royal Society, Ada Lovelace Institute. Learning data lessons: Data access and sharing during COVID-19 [document on the Internet]. c2021 [cited 2022 Mar 29]. Available from: https://royalsociety.org/-/media/policy/Publications/2021/learning-data-lessons-data-access-and-sharing-during-COVID-19.pdf?la=en-GB&hash=DA87DF3B44154E407FDADC6B4269CEED
18. Abebe R, Aruleba K, Birhane A, Kingsley S, Obaido G, Remy S, et al. Narratives and counternarratives on data sharing in Africa. In: Proceedings of the 2021 FAccT ACM Conference on Fairness, Accountability, and Transparency; 3–10 March 2021; virtual event. New York: Association for Computing Machinery (ACM); 2021. p. 329–341. https://doi.org/10.1145/3442188.3445897

19. Pyo S, Reggi L, Martin E. The potential role of open data in mitigating the COVID-19 pandemic: Challenges and opportunities. Health Affairs Blog. 2 November 2020 [cited 2022 Mar 29]. https://doi.org/10.1377/forefront.20201029.94898

20. Coronavirus Resource Center, Johns Hopkins University and Medicine. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [webpage on the Internet]. No date [cited 2022 Aug 23]. Available from: https://coronavirus.jhu.edu/map.html

21. The Royal Society. Science as an open enterprise. Science Policy Centre report 02/12. London: The Royal Society; 2012. Available from: https://royalsociety.org/~/media/policy/projects/sape/2012-06-20-saoe.pdf

22. Hedberg K, Maher J. Collecting data. In: Rasmussen S, Goodman R, editors. The CDC field epidemiology manual. New York: Oxford Academic; 2019. p. 53–70. https://doi.org/10.1093/oso/9780190933692.003.0004

23. World Health Organization (WHO). Integrated disease surveillance and response technical guidelines in the WHO Africa region. Booklet one: Introduction section. Brazzaville: WHO Regional Office for Africa; 2019. Available from: https://apps.who.int/iris/bitstream/handle/10665/325015/WHO-AF-WHE-CPI-05.2019-eng.pdf?sequence=1&isAllowed=y

24. Committee on Population; Division of Behavioral and Social Sciences and Education; US National Academies of Sciences, Engineering and Medicine. Sharing research data to improve public health in Africa: A workshop summary. O'Connell ME, Plewes TJ, rapporteurs. Washington DC: US National Academies Press; 2015. https://doi.org/10.17226/21801

25. Africa Centres for Disease Control and Prevention. Africa CDC event-based surveillance framework [document on the Internet]. c2018 [cited 2022 May 03]. Available from: https://africacdc.org/download/africa-cdc-event-based-surveillance-framework/

26. Africa Centres for Disease Control and Prevention. Public health information systems [webpage on the Internet]. No date [cited 2022 May 03]. Available from: https://africacdc.org/programme/public-health-information-systems/

27. Africa Centres for Disease Control and Prevention. Southern Africa RISLNET. World Bank-Africa CDC Regional Investment Financing Project [webpage on the Internet]. No date [cited 2022 May 03]. Available from: https://africacdc.org/southern-africa-rislnet/

28. Africa Centres for Disease Control and Prevention. Extension for Community Healthcare Outcomes (ECHO) [webpage on the Internet]. No date [cited 2022 May 03]. Available from: https://africacdc.org/programme/public-health-information-systems/extension-for-community-healthcare-outcomes-echo/

29. Africa Centres for Disease Control and Prevention. Digital disease surveillance [webpage on the Internet]. No date [cited 2022 May 03]. Available from: https://africacdc.org/programme/surveillance-disease-intelligence/digital-disease-surveillance/

30. Vryzakis A, Thereaux O. How our network is considering data ethics: Survey results. The Open Data Institue Blog. 11 March 2020. Available from: https://theodi.org/article/how-our-network-is-considering-data-ethics-survey-results/

31. Van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health. 2014;14, Art. #1144. https://doi.org/10.1186/1471-2458-14-1144

32. Lucas-Dominguez R, Alonso-Arroyo A, Vidal-Infer A, Aleixandre-Benavent R. The sharing of research data facing the COVID-19 pandemic. Scientometrics. 2021;126(6):4975–4990. https://doi.org/10.1007/s11192-021-03971-6

33. Moorthy V, Henao Restrepo AM, Preziosi MP, Swaminathan S. Data sharing for novel coronavirus (COVID-19). Bull World Health Organ. 2020;98(3):150. https://doi.org/10.2471/BLT.20.251561

34. Capocasa M, Anagnostou P, Bisol GD. A light in the dark: Open access to medical literature and the COVID-19 pandemic. Inf Res. 2022;27(2), Art. #929. https://doi.org/10.47989/irpaper929

35. UK Central Digital and Data Office. Data ethics framework [webpage on the Internet]. c2018 [cited 2022 Mar 29]. Available from: https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020

36. Mahomed S, Loots G, Staunton C. The role of data transfer agreements in ethically managing data sharing for research in South Africa. S Afr J Bioeth Law. 2022;15(1):26–30. https://doi.org/10.7196/SAJBL.2022.v15i1.807

37. Scimago Journal & Country Rank. Country rankings: Public Health, Environment, and Occupational Health [database on the Internet]. c2022 [cited 2022 Aug 23]. Available from: https://www.scimagojr.com/countryrank.php?region=Africa&category=2739

38. Suominen K, Vambell E. Alliance for E-Trade Development: Toward an African data transfer regime to enable MSMEs' cross-border ecommerce [document on the Internet]. c2021 [cited 2022 May 06]. Available from: https://www.allianceforetradedevelopment.org/_files/ugd/478c1a_72021e35a826441db0723642a79e65e5.pdf

39. The Parliament of the Republic of Ghana. Data Protection Act, 2012.

40. The Member States of the African Union. African Union Convention on Cyber Security and Personal Data Protection [document on the Internet]. c2014 [cited 2022 Aug 22]. Available from: https://au.int/sites/default/files/treaties/29560-treaty-0048_-_african_union_convention_on_cyber_security_and_personal_data_protection_e.pdf

41. Barchi F, Little MT. National ethics guidance in sub-Saharan Africa on the collection and use of human biological specimens: A systematic review. BMC Med Ethics. 2016;17(1), Art. #64. https://doi.org/10.1186/s12910-016-0146-9

42. Academy of Science of South Africa (ASSAf). Concept note: The Protection of Personal Information Act No 4 of 2013: A Code of Conduct for Research [document on the Internet]. c2021 [cited 2022 May 03]. Available from: https://www.assaf.org.za/files/2020/POPIA%20CoC%20Research_Conceptnote_Letter%20003.pdf

AUTHORS:
S'busiso M. Nkosi[1]
Nomaxhosa M. Msimango[2]

AFFILIATIONS:
[1]Technology Station in Chemicals, Mangosuthu University of Technology, Durban, South Africa
[2]Department of Chemistry, Mangosuthu University of Technology, Durban, South Africa

CORRESPONDENCE TO:
S'busiso Nkosi

EMAIL:
nkosis@mut.ac.za

DATES:
Received: 26 Aug. 2021
Revised: 14 Jan. 2022
Accepted: 03 Aug. 2022
Published: 30 Nov. 2022

HOW TO CITE:
Nkosi SM, Msimango NM. Screening of zinc, copper and iron in Chinese cabbage cultivated in Durban, South Africa, towards human health risk assessment. S Afr J Sci. 2022;118(11/12), Art. #12099. https://doi.org/10.17159/sajs.2022/12099

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☒ All data included
☐ On request from author(s)
☐ Not available
☐ Not applicable

EDITORS:
Priscilla Baker
Amanda-Lee Manicum

KEYWORDS:
copper, zinc, iron, health risk assessment, transfer factor

FUNDING:
Mangosuthu University of Technology, Technology Innovation Agency

Screening of zinc, copper and iron in lettuce and Chinese cabbage cultivated in Durban, South Africa, towards human health risk assessment

Human well-being and ecological reliability continue to face a major threat resulting from heavy metal pollution to soils caused by untreated discharge from metropolitan and industrial wastewater. The potential human health risks of zinc (Zn), copper (Cu) and iron (Fe) contamination to native inhabitants through the food chain were assessed in Pinetown, Durban, where their irrigation processes are from the Umgeni River passing through the highly industrialised Pinetown area. River water, vegetables (cabbage and lettuce) and soil were analysed for Zn, Cu and Fe; transfer factor, health risk index and the daily intake of metals were also calculated. The concentrations of heavy metals indicated the pattern trend as Fe>Zn>Cu for both cabbage and lettuce. The levels of transfer factors for heavy metals ranged from 0.02 mg/kg to 1.89 mg/kg. The health risk index (0.0002–01430) was found to be within the recommended range (<1), which poses no human health risk with respect to all heavy metals tested.

Significance:

The present study has generated data on heavy metal pollution in and around the area and associated risk assessment for consumers' exposure to the heavy metals. These data can assist decision-makers in understanding the suitability status of vegetable consumption and irrigation by providing an understanding of the human health risk of the studied area. This database can be used as a tool to pinpoint the mechanisms and processes influencing public health implications of heavy metals in foods, soils and water.

Introduction

Heavy metals are elements that occur naturally and are known to have a high density – more than five times that of water – and high atomic weight. Their wide distribution in the environment arises from their numerous applications such as agricultural, industrial, technological, domestic, and medical which raises worries over their potential effects on the environment and human well-being.[1] Copper (Cu), zinc (Zn), and iron (Fe) are some of the fundamental elements that are needed for different biological, chemical and physiological functions of plants. However, an insufficient supply of these elements in the environment causes syndromes, disorders, and deficiency diseases in human beings.[2] Wastewater systems of businesses and municipalities discharge heavy metals into the environment. These inappropriate actions result in contamination of water and soil, which are further used for farming, bringing about increased accumulation of heavy metals in vegetable plants and thereby affecting food security all through the world.[3] It has been documented that heavy metals can be taken up by vegetables and then accumulate in the edible parts of the plant[4], and, if the accumulation is sufficiently high, it could cause clinical issues for humans and other animals who consume these plants.[5] Hence, heavy metal contamination of vegetables cannot be overlooked because vegetables are a significant part of the human diet and their intake may be hazardous to human health and lead to various long-term diseases. Heavy metal sources include weathering of metal-bearing rocks and volcanic eruptions, while anthropogenic sources include mining and various industrial and agricultural activities.[5]

Copper is an essential nutrient involved in the creation of red blood cells, maintenance of the body defences and functioning of brain cells. Copper is found in soil and water within industrialised communities.[6] Consumption of high levels of copper can cause nausea, vomiting, diarrhoea, gastric (stomach) complaints and headaches. Long-term exposure over many months and years can cause liver damage and death.[6] Zn is considered a fundamental component for human existence; however, acute and chronic exposure to excessively high concentrations of Zn can cause nausea, vomiting, diarrhoea, fever and lethargy.[7] Fe is a widely distributed mineral that is vital for human, plant and animal life – it is necessary for the production of haemoglobin, myoglobin and certain enzymes. However, excess Fe in the system can cause cirrhosis when deposited in the pancreas, liver cancer when deposited in the liver and cardiac arrhythmias when deposited in the heart.[8]

An assessment of the risk to human health is essential as it informs the management stage of risks, including recommendations to ensure that human health is protected. The assessment, correction, regulation, monitoring, and protection of environmental attributes that can have a negative impact on human health, as well as the advancement of environmental consequences that can improve human health, are the focus of health and environmental systems.[9] A number of studies have demonstrated that polluted environments with heavy metals may cause certain vegetables, including cabbage and lettuce, to accumulate high concentrations of these metals.[10-13] Given the potential toxicity, recalcitrant nature, and cumulative behaviour of heavy metals; the frequency of vegetable consumption; and safety and health concerns, such research is essential. Primary objectives of this study were to:

(1) investigate the current level of local heavy metal pollution in vegetated soils, plants and water,

(2) assess potential health risks from heavy metals in the soil–vegetable and water systems, and

(3) serve as a resource on heavy metal pollution prevention and treatment for policymakers and decision-makers.

The present study was conducted on a local farm, the Fair Food Company & Edamame Development Programme in Pinetown, Durban, South Africa; with the aim of assessing Zn, Cu, and Fe concentrations in irrigation water, soil,

Chinese cabbage (*Brassica rapa pekinensis*), and leafy lettuce (*Lactuca sativa*) from two different sites on the farm, which uses river water for irrigation of agricultural land on a long-term basis. The concentrations of these heavy metals in the soil, vegetables, and water were also compared to recommended limits for human safety.

Materials and methods

Study area

The study area was located in Marianhill, Pinetown in KwaZulu-Natal, South Africa. Two sampling sites were chosen and were located at The Fair Food Company & Edamame Development Programme (Site 1: 29°50'27.9''S, 30°49'21.2''E; Site 2: 29°50'40.9''S, 30°49'25.9''S). River water from the Umgeni River was used for irrigation at both sites; the River passes through a highly popularised industrial area in Pinetown.

Water sampling

Samples of irrigation water were collected from a single location in the Umgeni River. Samples were collected into polythene bottles that were cleaned with metal-free soap, soaked in 2% nitric acid and then washed with demineralised water.[14] Samples were brought to the Analytical Services Laboratory, Technology Station in Chemicals, Mangosuthu University of Technology, Durban, and kept at 5 °C.

Soil sampling

Soil samples from the farm were collected by digging approximately 500 g of soil (0–30 cm) from four different areas (125 g) of each site using a plastic scoop. The samples were completely blended to form a uniform soil sample. Foreign particles such as grass and stones were removed from the samples and the soils were dried in the oven. Samples were filtered through a sieve (2 mm) and preserved in a labelled polythene bag.[15]

Vegetable sampling

The study area contained a range of vegetables; cabbage and lettuce were chosen for the study because they are sometimes eaten raw through salads. For each zone, 3–6 replicates of the entire vegetable parts (leaves, roots, and stems) were collected. Samples were placed in a properly marked sampling bag and taken to the laboratory. They were thoroughly washed under running tap water to eliminate aerial contaminants and soils and rinsed with distilled water. Roots, stems and leaves were separated and the samples were oven dried (55–60 °C), blended and sieved (40 mesh).[16]

Digestion of samples and analysis

Half a gram of each vegetable and soil sample was weighed into a Teflon vessel and digested by 12 mL of aqua regia, i.e. a mixture of nitric acid and hydrochloric acid in a 3:1 ratio. The vessels were closed and placed on the microwave digestion system (UniClever BM—1z, Plazmatronika, Poland). At the end of the digestion, a 20-min airing process (no microwave power) was undertaken in order to cool the vessels so as to reduce the pressure to within the ambient values. The samples were poured into 100-mL volumetric flasks and filled with distilled water. Whatman filter papers (No. 42) were used to filter the samples; a filtrate of 5 L per sample was stored at 5 °C. Heavy metal content was determined by a PerkinElmer 2380 Atomic Absorption Spectrophotometer fitted with a lamp specifically designed for specific heavy metals. The remainder of the parameters were the same as in the method described by Welz et al.[17]

Determination of soil and water pH

Soil pH was measured using a pH meter (soil solution ratio 1:2 in water); water pH was determined in situ using a portable pH meter (201T ATC) from Laboratory Equipment suppliers.

Determination of transfer factor

The bioaccumulation of metals in the environment is a highly dynamic process which relies upon explicit combinations of synthetic, natural and ecological conditions.[18] Heavy metals are taken up by plants from the soil through the roots, and also from the environment through overground vegetative organs.[19] Plants' requirements for micronutrients play a controlling role in metals availability for plants as well as their ability to assimilate and kill harmful elements. This accessibility is unique, and depends on the plant species and their adaptation to the climatic conditions. The soil-to-plant tissue transmission is investigated using the transfer factor (TF) index (Equation 1). TF is calculated as a ratio of the concentration of a specific metal in plant tissue (C_{plant}) to the concentration of the same metal in soil (C_{soil}), both represented in the same unit (mg/kg fresh weight).[20] Higher values of TF (≥ 1) translate to increased absorption of metals from soil to plants; meaning the phytoremediation and phyto-extraction will have higher suitability for the plant. A TF value lower than one (<1) indicates poor metal absorption by the plant and therefore no obvious risk will be observed.[21]

$$TF = C_{plant} / C_{soil} \quad \text{Equation 1}$$

Determination of daily intake of metals

Tolerable daily intake is an indication of the actual amount in nutrition or potable water that can be ingested on a body mass basis, commonly mg/kg body weight every day for a long period by people who do not face a significant health risk. The daily intake of metals (DIM) was determined from the edible part of the plants using Equation 2:

$$DIM = M \times K \times I/W, \quad \text{Equation 2}$$

where M is the concentration of metals in plants, K is the conversion factor, I is the daily intake of vegetables and W is average body mass. Average body mass for an adult and child were projected to be 59.9 kg and 32.7 kg, respectively.[22] Vegetable daily intake for adults and children was projected to be 0.345 and 0.232 kg/person/day, respectively. The fresh weight of vegetables was converted to the dry weight using a conversion factor of 0.085.

The health risk index (HRI) for ingestion of Cu, Fe, and Zn through polluted vegetables was calculated using Equation 3[23]:

$$HRI = DIM/Rfd \quad \text{Equation 3}$$

where the reference oral dose is represented by Rfd. Rfd values for Fe, Cu, and Zn are 0.70, 0.04, and 0.30 mg/kg bw/day, respectively[24].

Statistical analysis

Microsoft Excel was used to determine statistical data and the results are presented as the mean (±standard deviation).

Results and discussion

Hydrogen potential

Water and soil samples were also analysed for hydrogen potential (pH) as it has a significant impact on determining the suitability of water for irrigating vegetables. Optimal soil pH plays a major part in growing the best plants and vegetables. Adjusting soil pH or matching plants or vegetables to the soil pH is important; most plants grow between a pH of 4.5 to 8.0. Soil pH is important because acidity or alkalinity of the soil determines what plant supplements are available to plant roots. The best pH for cabbage and lettuce is 6.5–6.8 and 6.0–6.5, respectively, for ideal growth and to debilitate club root infections. The irrigation water was found to have a pH of 6.93, which is almost neutral, and soil samples from Sites 1 and 2 showed a pH of 5.55 and 6.35, respectively.

Concentrations of heavy metals in water and soil samples

Heavy metal concentrations in water and soil samples from Sites 1 and 2 are illustrated in Table 1. The content levels of Cu and Fe in water were measured to be 0.075 mg/kg and 0.731 mg/kg, respectively, which exceeds the WHO/FAO standard parameters of 0.017 mg/kg and 0.50 mg/kg, respectively. Zn levels in water were measured to be 0.131 mg/kg, which is in the range of the permissible limit of 0.20 mg/kg. All soil samples from Sites 1 and 2 were in the range of the permissible limits of the WHO/FAO: 0.285 mg/kg and 0.266 mg/kg for Cu, 89.87 mg/kg and 86.35 mg/kg for Fe, and 1.405 mg/kg and 1.361 mg/kg for Zn, respectively.

Table 1: Soil and water (mg/kg) heavy metal concentrations

Heavy metal	Soil		Water	WHO/FAO	
	Site 1	Site 2	Sites 1 and 2	Water	Soil
Cu	0.285 ± 0.020	0.266 ± 0.040	0.075 ± 0.001	0.017	100
Zn	1.405 ± 0.007	1.361 ± 0.017	0.131 ± 0.016	0.200	300
Fe	89.87 ± 0.460	86.35± 1.540	0.731 ± 0.050	0.500	50 000

Heavy metal concentrations in vegetables

The concentrations in mg/kg were assessed in roots, stem and leaves, as presented in Table 2 for both cabbage and lettuce grown at two different sites irrigated with river water. The mean concentrations of heavy metals in vegetables appeared to have a wide range: 0.061–0.339 mg/kg for Cu, 1.650–10.24 mg/kg for Fe and 0.387–2.566 mg/kg for Zn. The highest levels of Cu, Fe and Zn were recorded in lettuce leaves, cabbage roots and lettuce roots, respectively, from Site 2 for all. The lowest levels of Cu, Fe, and Zn were all recorded in cabbage leaves from Site 2. It is understood that the differences and variations in concentrations could be attributed to the plant's abilities in terms of absorbing and accumulating heavy metals[25], and differences in growth phase and rate of growth[26].

The contents of heavy metals studied were also compared to the permitted WHO/FAO maximum intake levels. Cu, Fe, and Zn concentrations in various parts of both vegetables were significantly lower than the maximum concentrations, meaning that the vegetables were safe for human consumption. The cabbage samples showed a trend for concentration to be in the order roots>stem>leaves, which may be due to the fact that roots transport nutrients, water content and heavy metals from the soil to the stem and then to the leaves. The results for cabbage followed a more or less similar trend to those of the study conducted by Meerkotter[27], in Cape Town, South Africa, in which the Cu concentrations in cabbage leaves were lower when compared to stems and roots; and the roots seemed to have a considerably higher concentration of Cu. In the same study, a similar trend was observed where cabbage roots showed greater concentrations of Fe and Zn than in either stems or leaves. The heavy metal concentrations for lettuce seemed to be portioned differently as there was no particular trend followed for Cu, Fe and Zn.

Transfer of heavy metals from soil to vegetables

The TF of trace elements from soil to plants is defined as the ratio of trace element contents in the plant (dry weight) to the total trace element contents in soil (dry weight).[28] A plant with the TF value greater or equal to one (≥1) is deemed more suitable for phytoextraction and phytoremediation, which is known to be due to the higher soil to plant metal absorption. Lower numbers, on the other hand, are an indication that those plants have a poor response to metal absorption and that the plant may be consumable by humans. Table 3 summarises the heavy metal transfer factor in the vegetables; The TF range for cabbage is 0.28–0.77 for Zn, 0.23–0.89 for Cu and 0.02–0.11 for Fe; for lettuce is 0.42–1.87 for Zn, 0.22–1.27 for Cu and 0.03–0.07 for Fe. The highest (1.27) and lowest (0.22) TF for Cu were found in lettuce leaves from both sites. The highest TF for Zn (1.87) was observed in lettuce roots and the lowest (0.28) in cabbage leaves, both from Site 2. As for Fe, the highest TF (0.11) was observed in cabbage roots from Site 1 and the lowest (0.02) was observed in cabbage leaves from Site 2. The TF values are comparable to the findings published by Jan et al.[29]

Table 3: Heavy metal transfer factors

Heavy metal	Plant part	Transfer factor plant/soil in vegetables			
		Cabbage		Lettuce	
		Site 1	Site 2	Site 1	Site 2
Cu	Roots	0.74±0.071	0.89±0.111	0.43±0.071	0.82±0.140
	Stem	0.45±0.013	0.38±0.243	0.39±0.054	0.68±0.021
	Leaves	0.45±0.012	0.23±0.031	0.22±0.006	1.27±0.043
Fe	Roots	0.11±0.020	0.05±0.053	0.05±0.041	0.05±0.008
	Stem	0.07±0.155	0.03±0.009	0.04±0.035	0.03±0.005
	Leaves	0.04±0.314	0.02±0.080	0.07±0.042	0.04±0.022
Zn	Roots	0.77±0.057	0.58±0.028	0.42±0.001	1.87±0.043
	Stem	0.64±0.023	0.37±0.065	0.46±0.044	0.45±0.020
	Leaves	0.62±0.033	0.28±0.025	0.89±0.040	0.49±0.101

Daily intake estimates and the health risk index of heavy metals for vegetables

The values in Table 4 represent data for DIM and HRI in adults and children. These values were calculated from the leaves as they are

Table 2: Heavy metal concentrations (mg/kg) in different parts of the vegetables

Heavy metal	Plant part	Mean concentration of heavy metals in vegetables				Permissible limits
		Cabbage		Lettuce		
		Site 1	Site 2	Site 1	Site 2	
Cu	Roots	0.211± 0.032	0.236± 0.050	0.123±0.002	0.218 ± 0.002	73
	Stem	0.129±0.011	0.102± 0.005	0.110±0.001	0.181 ± 0.018	
	Leaves	0.129± 0.042	0.061± 0.020	0.064±0.003	0.339 ± 0.003	
Fe	Roots	10.24±0.077	3.980± 0.038	4.159±0.026	4.734 ± 0.054	425
	Stem	6.038±0.311	2.890± 0.028	3.942±0.140	2.676 ± 0.075	
	Leaves	3.788±0.062	1.650± 0.008	6.038±0.311	2.985 ± 0.034	
Zn	Roots	1.077±0.279	0.789 ± 0.040	0.584±0.022	2.566 ± 0.042	100
	Stem	0.905±0.040	0.501± 0.006	0.641±0.015	0.611 ± 0.005	
	Leaves	0.874±0.022	0.387± 0.015	1.246±0.0110	0.666 ± 0.011	

Table 4: Daily intake of metals (kg/person/day) and health risk index in adults and children

Sites	Vegetables	Daily intake of metals			Health risk index		
		Cu	Fe	Zn	Cu	Fe	Zn
Adults							
Site 1	Cabbage	0.0057±0.014	0.0018±0.035	0.0004±0.004	0.1430±0.024	0.0026±0.018	0.0014±0.025
	Lettuce	0.0003±0.010	0.0029±0.051	0.0006±0.009	0.0076±0.031	0.0041±0.023	0.0019±0.042
Site 2	Cabbage	0.0003±0.022	0.0008±0.023	0.0002±0.015	0.0008±0.033	0.0011±0.088	0.0006±0.005
	Lettuce	0.0002±0.011	0.0014±0.021	0.0003±0.014	0.0040±0.030	0.0020±0.027	0.0011±0.036
Children							
Site 1	Cabbage	0.0009±0.019	0.0289±0.024	0.0062±0.008	0.0230±0.036	0.0413±0.014	0.0207±0.021
	Lettuce	0.0005±0.032	0.0429±0.026	0.0089±0.003	0.0113±0.029	0.0612±0.024	0.0295±0.073
Site 2	Cabbage	0.0004±0.053	0.0117±0.020	0.0028±0.016	0.0108±0.046	0.0167±0.029	0.0092±0.025
	Lettuce	0.0024±0.005	0.0212±0.053	0.0047±0.043	0.0603±0.072	0.0303±0.006	0.0158±0.064

the most edible parts of the vegetables. The estimated dietary intakes for metals were below the tolerable limits. The results indicate that the highest intake in adults was 0.0057 kg Cu/person/day in cabbage from Site 1; and the highest intake in children was 0.0429 kg Fe/person/day in lettuce from Site 1. RfD for heavy metals is known to be 0.04 mg/kg/day for Cu, 0.7 mg/kg/day for Fe and 0.3 mg/kg/day for Zn.[30] The RfD is seen as an estimate of day-to-day accessibility to the general public that is unlikely to present a tremendous risk of detrimental consequences over the course of a lifetime.[30] The HRI results were obtained by dividing daily intake of heavy metals by their reference measurements to evaluate the health risk that comes with these heavy metals (Table 4). An HRI less than 1 demonstrates that the assessment is unlikely to pose a significant health risk. However, a HRI greater than 1 does not imply that a serious adverse health impact will emerge, it just indicates a strong probability of a health risk. The HRI of the study area suggests that cabbage and lettuce grown at both sites of The Fair Food Company & Edamame Development Programme were totally free from any risk and safe for consumption. Despite the fact that all of the heavy metals in the samples examined were within the global limit for agricultural use, long-term reuse of irrigated water causes an excessive build-up of those hazardous metals in soil and crops. Hence, proper waste management and environmental practices in the surrounding areas are critical.

Conclusion

The concentration of various elements in plants is determined by the approximate degree of plant interaction with polluted soil as well as hazardous element deposition in air pollutants. Human health risk assessment helps experts to assess the overall situation and determine what advice or actions, if any, should be taken to ensure that human health is protected. Past, current or future exposures to heavy metals in air, soil, water, food, consumer products or other materials can be assessed.

Although the river water for irrigation passes through a highly industrial urban area, our results reveal that the vegetables were least contaminated by Cu, Fe and Zn, which indicates less discharge of these metals into the natural ecosystem by nearby industries. It is recommended that, to keep the environment less affected by heavy metals, proactive health agencies, trash disposal knowledge, and best practices should be maintained. Because the two farming sites share a single source of irrigation water, the overall differences in heavy metal concentrations were minimal and insignificant. It is noteworthy to mention that soil concentrations from the two selected sites were also not significantly different.

This study reports that the human assessment risk for selected metals is low, hence vegetables may be considered safe for consumption. However, some data in this study are lacking as the first screening focus was mainly on three heavy metals. A more detailed study is required to investigate a wider range of heavy metals in the area and surrounding areas. The outcome of this investigation could be used as a tool for farmers and decision-makers to adopt and implement action-oriented, sustainable strategies to prevent risk to the population from heavy metals in vegetables.

Acknowledgements

We acknowledge funding from Technology Station in Chemicals and the Department of Chemistry, Mangosuthu University of Technology, Durban, South Africa, and the Technology Innovation Agency (TIA).

Competing interests

We have no competing interests to declare.

Authors' contributions

S.M.N: Conceptualisation, methodology, sample analysis, drafted the first manuscript, project leadership. N.M.M: Sample analysis, data collection, manuscript revision.

References

1. Lazur A, Van Derwerker T, Koepenick K. Review of implications of road salt use on groundwater quality-corrosivity and mobilization of heavy metals and radionuclides. Water Air Soil Pollut. 2020;231:474–500. https://doi.org/10.1007/s11270-020-04843-0
2. WHO/FAO/IAEA. Trace elements in human nutrition and health. Geneva: World Health Organization; 1996.
3. Ratnakar S, Bhabagrahi S. A simplified modelling framework for real-time assessment of conservative pollutants in ungauged rivers during cloudy periods. J Environ Manag. 2021;293:11282–11289. https://doi.org/10.1016/j.jenvman.2021.112821
4. Ram P, Tapos KM, Saiful I, Krishno C. Potential health risk of heavy metals via consumption of rice and vegetables grown in the industrial areas of Bangladesh. Hum Ecol Risk Assess. 2020;26:921–943. https://doi.org/10.1080/10807039.2018.1546114
5. Islam MM, Avha NJ, Ahmed S. Trace metals and organochlorine pesticide residues in imported fishes in Bangladesh and human health risk implications. Environ Sci Pollut Res. 2021;14:258–261. https://doi.org/10.1007/s11356-021-17074-5
6. Zwolak A, Sarzyńska M, Szpyrka E. Sources of soil pollution by heavy metals and their accumulation in vegetables: A review. Water Air Soil Pollut. 2019;230:164–170. https://doi.org/10.1007/s11270-019-4221-y

7. Jayaweera L, Shafras M, Navaratne A, Sanas M, Antonypillai C. Serum zinc levels in diabetic patients with and without dyslipidaemia and cardiovascular diseases. Sri Lanka J Diabetes Endocrinol Metab. 2020;10(2):3–9. http://doi.org/10.4038/sjdem.v10i2.7414

8. Nagla RE, Michael TT, Greenwood T. Iron mediated toxicity and programmed cell death: A review and a re-examination of existing paradigms. Biochim Biophys Acta Mol Cell Res. 2017;1864(2):399–430. https://doi.org/10.1016/j.bbamcr.2016.12.002

9. García-Carmona MA, Romero-Freire M, Sierra Aragón F, Martín Peinado J. Effectiveness of ecotoxicological tests in relation to physicochemical properties of Zn and Cu polluted Mediterranean soils. Geoderma. 2019;107:259–268. https://doi.org/10.1016/j.geoderma.2018.12.016

10. Danjuma MS, Abdulkadir B. Bioaccumulation of heavy metals by leafy vegetables grown with industrial effluents: A review. Bayero J Pure Appl Sci. 2018;29:934–939. https://doi.org/10.4314/bajopas.v11i2.23

11. Mohsen J, Javad KM. Assessment of the health risks of heavy metals in soils and vegetables from greenhouse production systems in Iran. Int J Phytoremediation. 2020;22:834–848. https://doi.org/10.1080/15226514.2020.1715917

12. Solankey SS, Kumari M, Akhtar S, Singh HK, Ray PK. Challenges and opportunities in vegetable production in changing climate: Mitigation and adaptation strategies. In: Solankey SS, Kumari M, Kumar M, editors. Advances in research on vegetable production under a changing climate. Vol. 1: Advances in olericulture. Cham: Springer; 2021. p. 17–25. https://doi.org/10.1007/978-3-030-63497-1_2

13. Tea MR, Andrea Č, Abra P, Jelena D, Jovica L, Dario O, et al. Physiological and morphological response of marine diatom *Cylindrotheca closterium* (Bacillariophyceae) exposed to cadmium. Eur J Phycol. 2021;56:24–36. https://doi.org/10.1080/09670262.2020.1758347

14. Elvira MV, Faustino-Eslava DV, De Chavez ERC. Human health risk associated with heavy metals from consumption of Asiatic clam, *Corbicula fluminea*, from Laguna de Bay, Philippines. Environ Sci Pollut Res. 2021;28:36626–36639. https://doi.org/10.1007/s11356-021-13298-7

15. Lei M, Liao B, Zeng Q, Qin P, Khan S. Fraction distribution of lead, cadmium, copper, and zinc in metal contaminated soil before and after extraction with disodium ethylenediaminetetraacetic acid. Commun Soil Sci Plant Anal. 2008;39(13–14):1963–1978. https://doi.org/10.1080/00103620802134776

16. Jamali MK, Kazi TG, Arain MB, Afridi HI, Jalbani N, Kandhro GA, et al. Heavy metal accumulation in different varieties of wheat (*Triticum aestivum* L.) grown in soil amended with domestic sewage sludge. J Hazard Mater. 2009;164:1386–1391. https://doi.org/10.1016/j.jhazmat.2008.09.056

17. Welz B, Sperling M. Atomic absorption spectrometry. 3rd ed. Weinheim: Wiley; 1999. https://doi.org/10.1002/9783527611690

18. Somdeep G, Madhurima B, Shouvik M, Punarbasu C, Assessment of role of rhizosphere process in bioaccumulation of heavy metals in fine nutritive roots of riparian mangrove species in river Hooghly: Implications to global anthropogenic environmental changes. Mar Pollut Bull. 2022;174, Art. #113157. https://doi.org/10.1016/j.marpolbul.2021.113157

19. Štofejová L, Juraj F, Danica F. Analysis of heavy metal content in soil and plants in the dumping ground of magnesite mining factory Jelšava-Lubeník (Slovakia). Sustainability. 2021;13:4508–4520. https://doi.org/10.3390/su13084508

20. Panuccio MR, Sorgonà A, Rizzo M, Cacco G. Cadmium adsorption on vermiculite, zeolite and pumice: Batch experimental studies. J Environ Manage. 2009;90:64–374. https://doi.org/10.1016/j.jenvman.2007.10.005

21. Al-Taani AA, Nazzal Y, Howari FM, Iqbal J, Bou Orm N, Xavier CM, et al. Contamination assessment of heavy metals in agricultural soil, in the Liwa Area (UAE). Toxics. 2021;9:53–70. https://doi.org/10.3390/toxics9030053

22. Latif A, Bilal M, Asghar W, Azeem M, Ahmad MI, Abbas A, et al. Heavy metal accumulation in vegetables and assessment of their potential health risk. J Environ Anal Chem. 2018;5, Art. #1000234. https://doi.org/10.4172/2380-2391.1000234

23. US Environmental Protection Agency. Integrated Risk Information System [homepage on the Internet]. c2018. Available from: https://www.epa.gov/iris

24. Rangnekar SS, Sahu SK, Pandit GG, Gaikwad VB. Study of uptake of Pb and Cd by three nutritionally important Indian vegetables grown in artificially contaminated soils of Mumbai, India. Int Res J Environ Sci. 2013;2:1–5.

25. Shimod KP, Vineethkumar V, Prasad TK. Effect of urbanization on heavy metal contamination: A study on major townships of Kannur District in Kerala, India. Bull Natl Res Cent. 2022;46:4–21. https://doi.org/10.1186/s42269-021-00691-y

26. Mina T, Azadeh K. Potential health risk of heavy metals accumulation in cultivated mulberry in urban landscapes of Arak, Iran: A case study. Arch Hyg Sci. 2019;8:274–285. https://doi.org/10.1016/0048-9697(92)90200-C

27. Meerkotter M. Sources of heavy metals in vegetables in Cape Town, and possible methods of remediation [PhD thesis]. Cape Town: University of the Western Cape; 2012.

28. Ahmad K, Wajid K, Khan ZI. Evaluation of potential toxic metals accumulation in wheat irrigated with wastewater. Bull Environ Contam Toxicol. 2019;102:822-828. https://doi.org/10.1007/s00128-019-02605-1

29. Jan FA, Ishaq M, Khan S, Ihsanullah I, Ahmad I, Shakirullah M. A comparative study of human health risks via consumption of food crops grown on wastewater irrigated soil (Peshawar) and relatively clean water irrigated soil (lower Dir). J Hazard Mater. 2010;179:612–621. https://doi.org/10.1016/j.jhazmat.2010.03.047

30. US Environmental Protection Agency. Integrated Risk Information System [webpage on the Internet]. c2010 [cited 2016 May 04]. Available from: http://www.epa.gov/iris/index.html

AUTHORS:
Enobong R. Essien[1]
Violette N. Atasie[1]
Samson T. Scott[1]
Opeyemi A. Ajayi[1]

AFFILIATION:
[1]Department of Chemical and Food Sciences, Bells University of Technology, Ota, Ogun State, Nigeria

CORRESPONDENCE TO:
Enobong Essien

EMAIL:
reggiessien@gmail.com

DATES:
Received: 14 Oct. 2021
Revised: 30 July 2022
Accepted: 02 Aug. 2022
Published: 30 Nov. 2022

HOW TO CITE:
Essien ER, Atasie VN, Scott ST, Ajayi OA. A tetracycline hydrochloride-loaded SiO_2/polycaprolactone composite from bamboo stem for controlled drug release study. S Afr J Sci. 2022;118(11/12), Art. #12560. https://doi.org/10.17159/sajs.2022/12560

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☒ All data included
☐ On request from author(s)
☐ Not available
☐ Not applicable

EDITORS:
Priscilla Baker
Amanda-Lee Manicum

KEYWORDS:
bamboo stem, polycaprolactone, composites, controlled drug release, silica, tetracycline hydrochloride

FUNDING:
None

A tetracycline hydrochloride-loaded SiO_2/polycaprolactone composite from bamboo stem for controlled drug release study

A controlled drug delivery system is preferable to traditional drug administration because it can supply the drug continuously and ensure on-demand bioavailability. The production of silica/polymer composite delivery material is expensive due to the use of alkoxysilane silica precursors. As bamboo is an abundant plant in Africa, we investigated the use of bamboo stems as an alternative silica starting material. The ash from the bamboo stem was mixed with polycaprolactone (PCL) solution to produce a (SiO_2/PCL) composite, which was then loaded with the drug, tetracycline hydrochloride (TCH), to test in vitro degradability and controlled-release in phosphate-buffered saline (PBS). Scanning electron microscopy, X-ray diffractometry and Fourier transform infrared spectroscopy were used to examine the structure, phase composition, and chemical bond properties of the material. The TCH release profile was determined using an ultraviolet (UV) spectrophotometer. The SiO_2/PCL composite showed a high capacity for drug loading. The composite released TCH in a consistent and sustained way, and showed regulated degradability in PBS. As a result, the use of bamboo stem-derived silica in the formulation of SiO_2/PCL for continuous TCH delivery shows considerable cost-benefit potential for a safe, regulated drug delivery strategy.

Significance:

- This study shows the benefit of using bamboo stem as an alternative silica source to alkoxysilanes.
- SiO_2/PCL composites can be employed for the sustained delivery of drugs while providing congruent degradation.
- This study can serve as a benchmark for further utilisation of bamboo stem as a low-cost silica precursor.

Introduction

A drug delivery system is a formulation or material capable of injecting a medication into the body while managing the speed, duration, and location of its release to preserve its effectiveness and safety. Drug delivery systems are important in drug pharmacology because they can impact drug release and speed, as well as drug distribution inside the body and reduce the occurrence of side effects. A good drug delivery system guarantees that an active drug is released at the proper moment and exerts its effects while still in the body.[1] Only a small portion of the drug reaches the right spot in the body during normal use, and the majority of it is eliminated through metabolism and enzyme secretion. Each drug has a therapeutic range in terms of concentration, with the greatest safe concentration being toxic and the lowest effective concentration being ineffective.[2]

Wen and Park[2] point out that a more regulated or zero-order release from microcontainers is desirable because controlling the drug release rate has several benefits, including boosting therapeutic impact, decreasing side effects, and reducing administration frequency. A system with continuous drug release must be biodegradable and biocompatible to release a therapeutic agent into the body and enhance the degree of effectiveness and safety of the treatment by modifying the rate, time, and place of release.

The drug can be embedded in a polymer matrix and released in a regulated manner due to the presence of the polymer matrix. Poly(lactide-co-glycolide), polylactide, polyesters, polyglycolide, and poly(hydroxybutyrate) have all been explored extensively as materials for drug administration.[3,4]

Polycaprolactone (PCL) is a synthetic polyester that has a low glass transition temperature (T_g) of about 60 °C. PCL has a high permeability due to its semi-crystallinity caused by its low T_g, which makes it useful for delivering drugs with a low molecular weight. PCL degrades at a slower rate than other biocompatible polymers like poly(L-lactide) and poly(lactide-co-glycolide) due to its semi-crystalline form. As a result, drug diffusion often dictates drug release from PCL matrices, making PCL ideal for regulated and sustained drug release. However, because drug release from PCL is primarily driven by diffusion, drug distribution within the polymer matrix has a significant impact on drug release rate. The influence of drug characteristics and polymer composition on drug release from PCL microspheres and thin films has been explored in previous work by Wang et al.[5] and Schlesinger et al.[6]

PCL has a low biodegradation rate, but good biocompatibility and high steroid permeability[7], which explains its use in an implanted, 1-year contraceptive delivery system for levonorgestrel[8]. Progesterone (Mw 314Da) has been successfully loaded into microporous PCL matrices by co-dissolving the steroid with PCL in acetone before precipitation casting.[9] In phosphate-buffered saline (PBS), high drug loadings of progesterone/mg matrix/day were possible after 11 days. Furthermore, anti-proliferative action of the steroid on breast cancer cells demonstrated that it was released with preserved activity. To include hydrophilic drug species like gentamicin sulfate into microporous PCL matrices made by precipitation casting, the drug powder must first be dispersed in a PCL solution. This formulation technique may sustain gentamicin sulfate release for 11 weeks in vitro, revealing the development of an interconnected pore network within the matrix that allows for efficient diffusion of low molecular weight water-soluble species.[10]

PCL has previously been found to have good ocular tolerability in the form of nanocapsule drug carriers for carteolol chlorhydrate.[11] This finding, combined with the ability of microporous PCL matrices to sustain gentamicin

sulfate release and the inherent flexibility of the PCL carrier (5–10 N/m^2 compressive strength)[12,13], suggests that highly water-soluble antibiotics such as tetracycline antibiotics could be delivered topically.

Furthermore, porous silica particles as an encapsulation material for drugs offer an interesting advantage in drug delivery due to their small size, ability to manage surface charge, optical properties, high surface area, low density and adsorption capacity delivery.[13,14]

Several biomasses have been investigated as silica sources based on their renewability and eco-friendliness.[15-17] The bamboo plant can thrive in any kind of climate and in very poor soils. Besides its renewability, the preponderance of bamboo, which is one of the most productive and fastest-growing natural resources[18], makes it an attractive natural resource. In Africa where it is widely available, it is regarded as 'a potential green gold'[19], implying that it could offer huge economic benefit if used as a substitute silica precursor to expensive alkoxysilanes on a large scale. Therefore, in this study, we aimed to investigate the use of silica derived from bamboo stem for the development of a controlled drug delivery system based on polycaprolactone.

Materials and methods

Materials

Bamboo (*Bambusa vulgaris*) stems (Figure 1) were obtained from a swampy area at the University of Lagos in Lagos, South-West, Nigeria. Authentication of the sample was performed by the Herbarium at the University of Lagos (number: LUH5493). The analytical grade reagents used for this experiment which were purchased from Sigma-Aldrich included: polycaprolactone (PCL, average molecular weight = 80 000); chloroform (HPLC grade, assay 99%); tetracycline hydrochloride (TCH, 95%); double-distilled de-ionised water, sodium chloride (98%), sodium dihydrogen phosphate (99%), potassium chloride (99%) and disodium hydrogen phosphate (99%).

Figure 1: Bamboo stem.

Biosilica preparation

The method used to extract the biosilica from the bamboo stem was similar to the approach by Kow et al.[20] Accordingly, the stem was cut into pieces, washed with double-distilled de-ionised water and sun-dried. Afterwards, the stem pieces were subjected to further drying in a hot air oven at 105 °C for 72 h and then ground to form a powder to increase the surface area. The powdered material was transferred into a muffle furnace and burned at 800 °C for 6 h at a heating rate of 10 °C/min to form the biosilica.

SiO_2/PCL composite preparation

To prepare the SiO_2/PCL composite via solvent casting, 0.5 g PCL powder was dissolved in chloroform (50 mL) by sonication (ultrasonic cleaner, CLEAN 120HD) at 40 W, 30 °C for 5 min. Following this, the as-prepared biosilica (1.0 g) was dispersed in the solution and the temperature was adjusted to 45 °C while the mixture was sonicated for another 2 h to give a uniform colloidal solution. After completing this procedure, the mixture was immediately cast into five cylindrical moulds of dimension 12 mm × 6 mm (height x diameter) and kept to dry at ambient condition for 5 days. The sample was tagged SiPCL after removal from the mould.

SiO_2/PCL-TCH composite preparation

To load the TCH onto the SiO_2/PCL, the procedure for obtaining the SiO_2-PCL mixture was repeated with a similar concentration. However, 2 min after adding the biosilica, 150 mg of TCH was dispersed in the mixture and sonicated at 45 °C for 2 h to give a homogeneous solution. The obtained mixture was then cast into five cylindrical moulds measuring 12 mm × 6 mm (height x diameter) and allowed to dry at room temperature for 5 days. When the drying process was completed, the samples were removed from the mould and labelled SiPCL/TCH.

Characterisation

The drug-loaded and unloaded composites, in quintuplicate, were assessed based on their tensile and compressive strengths by a universal testing machine (MODEL BAB-200, Transcell Technology) with a 100 kgf load, at a crosshead rate of 50 mm/min. The test was conducted under a temperature of 24±1 °C and relative humidity of 45±5% in accordance with ASTM standard D5034-95. The dimensions of the samples were 22 mm in nominal length and 5 mm in width based on ASTM standard D1708-96. With the aid of Instron pneumatic clamps, the tensile tests were performed at a strain rate of 0.01/s from where the load and displacement data were obtained. The data were analysed using a two-way analysis of variance (ANOVA; Prism, GraphPad).

The elemental composition of the biosilica was determined in an energy dispersive X-ray fluorescence (EDXRF) spectrometer (Minipal 4). The samples were converted into fine homogeneous particles by sieving through a 150-micron mesh sieve. The voltage used for oxides was 14 kV with a Kapton filter, while for trace elements an Ag/Al thin-film filter was set at 20 kV.

To examine the microstructure of the composites, a scanning electron microscope was used (Phenom ProX, Phenomworld). The samples were distributed over carbon conductive tape and sputter-coated with 5 nm gold using a quorum-Q150R Plus E sputter coater.

For phase composition and crystallinity assessment, an X-ray diffractometry (XRD; Empyrean Panalytical) system, using Cu-Kα radiation with a wavelength of 1.5418 nm and a tube current of 40 mA to generate the X-ray, was employed. The samples were scanned through the diffraction angle from 4° to 75° with a two-step of 0.026261 at 8.67 s for each step.

A Fourier transform infrared spectrometer (FTIR; Agilent Technology Cary 630) in transmittance mode with attenuated total reflectance was used to obtain the spectra depicting the chemical bond properties of the composites. To obtain the spectra, an average of 30 scans at 8 cm^{-1} resolutions were performed over the spectral scan range of 4000–650 cm^{-1}.

In vitro drug release experiment

To study the in vitro drug release, the SiPCL/TCH composite (0.25 g) was immersed in 100 mL PBS release medium at a pH of 7.4 in a clear glass vial. Afterwards, it was incubated at 37 °C with constant agitation at 100 rpm for a maximum of 28 days. During the incubation period, the PBS solution was withdrawn at pre-defined time intervals to measure the absorbance with a UV/visible spectrophotometer (T70+ UV/Vis Spectrometer, PG Instruments Ltd.) at a wavelength of 358 nm to determine the concentration of TCH released from the composite, in accordance with previous studies.[21,22] The TCH stock concentration of 5−100 μg/mL in PBS solution was initially used to establish the calibration curve from which the drug concentration was extrapolated using an absorbance-concentration calibration curve. The cumulative TCH released from the SiPCL/TCH composite was computed relative to

the original drug weight in the polymer matrix and plotted against time. The drug release testing was carried out in triplicate.

In vitro degradation test

The weight loss of the SiPC/TCH composite was investigated for a maximum of 28 days by incubating 0.5 g of the sample in 30 mL of PBS (pH 7.4) at 37 °C with continual stirring at 100 rpm. The samples were kept in glass vials with a cover to prevent the PBS buffer from evaporating. After removal from PBS at intervals of 7, 14, 21 and 28 days, the samples were washed with double-distilled de-ionised water and dried at 105 °C until they reached a consistent weight, then cooled and kept in a desiccator at room temperature for 60 min. The degradation (weight loss) in % was calculated according to Equation 1.[23] The dry weights of the original and degraded specimens are W_0 and W_t, respectively, and the degradation rate is D.

$$D = [(W_0 - W_t)/W_0] \times 100\% \quad \text{Equation 1}$$

Statistical analysis

One-way ANOVA was used to determine the difference between the drug cumulative amounts released from the tested samples at each immersion time; values are mean±standard deviation. This analysis was followed by an appropriate post-hoc test, where $p<0.05$ was taken as a criterion for a statistically significant difference.

Results and discussion

Composition of the bamboo stem-derived silica

The silica production was 37.16% based on the dry weight of the bamboo leaf biomass. The elemental composition of the calcined bamboo leaves as determined by EDXRF is given in Table 1. The analysis shows that the abundance of silicon is 58.31%. This finding contrasts with the study of Dirna et al.[24] who obtained 42.84% silicon from bamboo stem after burning it at 700 °C for 6 h in a muffle furnace. From the result presented, SiO_2 was present in the bamboo stem up to 93.58%. No heavy metals were found, thus indicating that the silica is safe for medical use.

Table 1: Elemental composition of the bamboo stem-derived silica as determined by X-ray fluorescence

Element	Atomic %
Si	58.31
O	35.27
K	3.48
Mg	1.74
S	1.13
Ni	0.07
Total	**100**

Mechanical properties

The mechanical properties assessment of SiPCL gave 1.4462±0.31 MPa as the compressive strength and 8.327±0.43 MPa as the tensile strength. Based on the results of the stress-strain curves of typical composite scaffolds, these low values could be attributed to the porous nature of the microstructure. These pore structures could enable the encapsulation and diffusion of drugs if applied in controlled drug release. For the SiPCL/TCH, the compressive strength was 1.236±0.22 MPa whereas the tensile strength was 7.776±0.38 MPa. Previous studies[25,26] have shown that, at certain thresholds, drug loading could cause a reduction in mechanical properties. This finding has been attributed to a decrease in the polymer crystallinity[25] due to the impregnation of the drug into the polymer matrix.

Morphology

The microstructure of SiPCL (Figure 2) demonstrates the presence of SiO_2 crystal aggregates contained in the polymer matrix. These spots can be found everywhere throughout the surface of the polymer matrix. The surfaces where the particles are embedded appear flat and broad, resulting in a vast surface area. The surface became more compact with a larger number of particle aggregates, an increase in the number of embedment sites, and a substantially bigger surface area after loading with TCH (SiPCL/TCH) (Figure 3). A large surface area is necessary to provide the high specific capacity required to establish an enhanced drug diffusion gradient in the composite.[27]

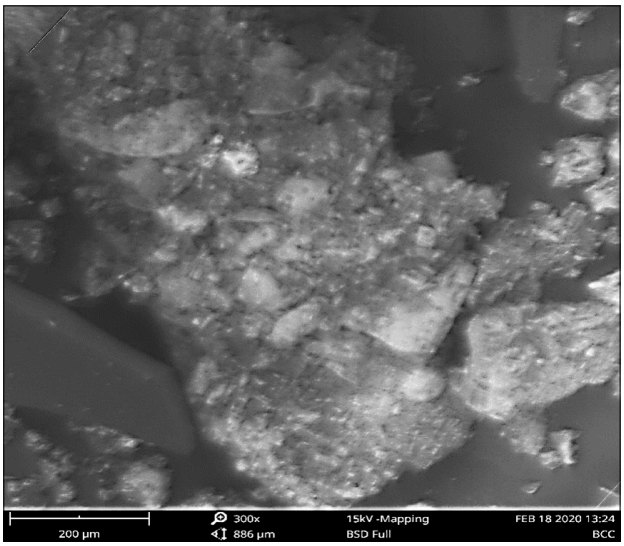


Figure 2: Scanning electron micrograph showing the microstructure of the SiPCL composite.

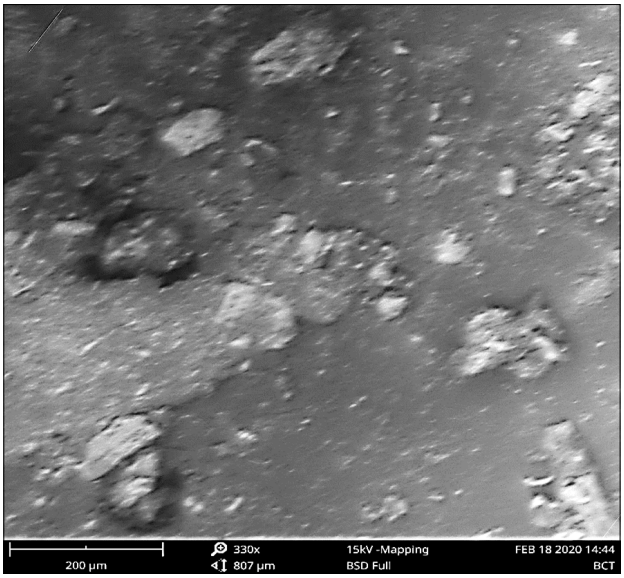


Figure 3: Scanning electron micrograph showing the microstructure of the composite loaded with drug (SiPCL/TCH).

Diffraction patterns

Figure 4 shows the diffractograms that illustrate the phase identification of the crystallographic structure of the composites. Peaks in the SiPCL diffraction pattern match both the PCL and SiO_2 components in the sample. When indexed with the standard reference file ICDD # -01-077-1317[28], prominent peaks at 2θ 21.5°, 26.5°, 28.5°, 31.5°, 36.5°, 39.7°, and 60° match cristobalite (SiO_2) in both angular position and intensity; those corresponding to PCL are situated at 2θ 21.3° and 23.6°.[29]

The appearance of the cristobalite peaks at high intensity validates the X-ray fluorescence result, which shows a high content of SiO_2 in the bamboo stem ash. The sharpness and intensity of the diffraction peaks, as well as the nature of their pattern at the baseline, indicate that the SiPCL sample is semi-crystalline.

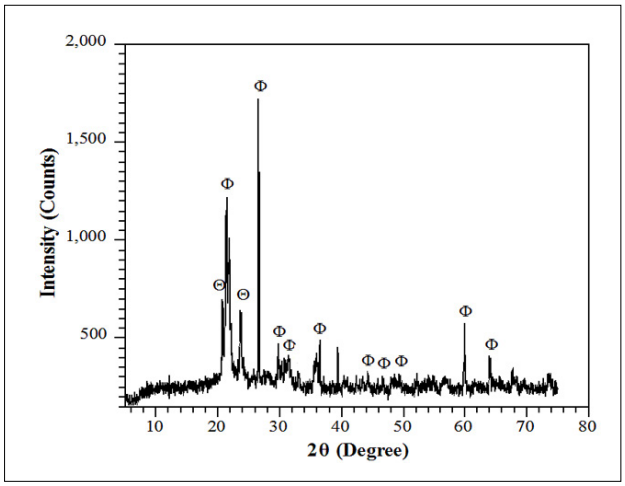


Figure 4: X-ray diffraction pattern of SiPCL composite showing the presence of two major phases: cristobalite (Φ) and PCL (Θ).

Figure 5 shows the XRD pattern after drug loading of the SiPCL composite (SiPCL/TCH). All SiO_2 and PCL peaks that were found in the SiPCL spectra are still present. In addition, two additional peaks corresponding to TCH crystal plane reflections appear at angular coordinates 28.69° and 42.47°.[30] In the spectrum for SiPCL/TCH, there was a small drop in baseline intensity. This is due to impregnation of the drug into the polymer matrix[31], and hence, supports the reduction in the mechanical strength after TCH loading of the composite discussed earlier.

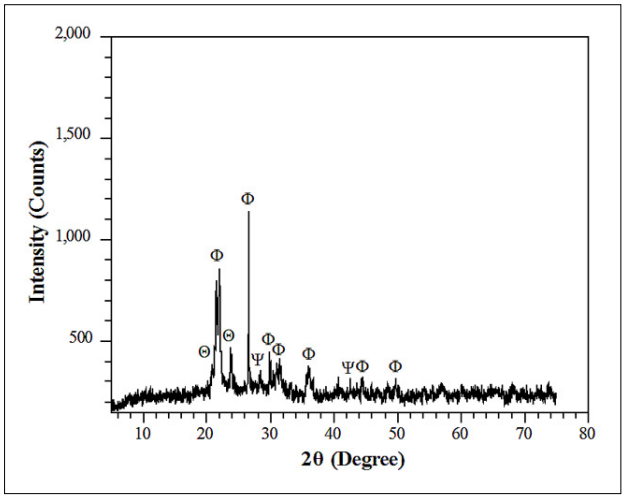


Figure 5: X-ray diffraction pattern of SiPCL/TCH composite showing the presence of TCH in the composite: cristobalite (Φ), PCL (Θ) and TCH (Ψ).

Chemical bond characteristics

The vibrational modes of the sample without the drug (SiPCL) are shown in the FTIR spectrum presented in Figure 6. The spectrum contains bands at 1428 cm^{-1} and 1368 cm^{-1} with low intensity considered for the methylene (CH_2) groups in amorphous regions in the PCL.[32] The small peak near 1238 cm^{-1} is attributed to the asymmetric vibrational stretching of C–O–C bonds of ester, which is further supported by an intense band that splits into two sharp peaks at 1001 cm^{-1} and 988 cm^{-1}, corresponding to the bending mode of C–O–C bonds in ester.[32] The sharp nature of these peaks signifies the development of crystallinity. The presence of PCL in the sample is further indicated by the C–H stretching vibration at 2937 cm^{-1}.

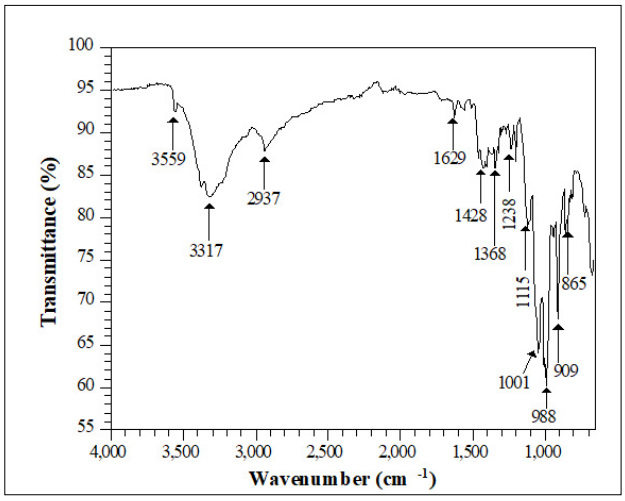


Figure 6: Fourier transform infrared spectrum of SiPCL showing the vibrational modes of bonds related to SiO_2 and PCL.

The characteristic vibrational stretching mode of the carbonyl (C=O) ester bond in PCL usually observed around 1720 cm^{-1},[33] could not be found in the spectrum. This may be due to its involvement in a crosslinking reaction with an oxygen atom of the SiO_2 via nucleophilic attack on the PCL carbonyl to form a network structure. Accordingly, two O–H peaks at 3559 cm^{-1} and 3317 cm^{-1} are observed in the spectrum, which are considered for the stretching vibrations of O–H group in the PCL and that result from the crosslinking reaction. The peak at 1115 cm^{-1} corresponds to the asymmetric stretching vibration of Si–O–Si and the ones at 909 cm^{-1} and 865 cm^{-1} are ascribed to symmetric stretching vibrations of Si–O–Si bonds in the silica and hence confirm the successful incorporation of silica into the polymer.[34]

The hydrophilicity of the silica was demonstrated by the angular vibration surface hydroxyl in the sample observed around 1629 cm^{-1}.[35] The appearance of the hydroxyl group can be attributed to the silanol bonds solely responsible for the adsorptive behaviour of hydrophilic silica and silicate surfaces.

After drug loading, the sample (SiPC/TCH) gave some shifts in vibrational frequencies as observed in Figure 7. The C–O–C ester bond stretching became visible around 1245 cm^{-1} while the twin sharp bending modes fused to develop a band around 1029 cm^{-1}. Several bands around that region disappeared, including the symmetric stretching Si–O–Si vibrational peaks around 909 cm^{-1} and 865 cm^{-1}; while the asymmetric stretching vibrational peak observed at 1001 cm^{-1} in SiPCL collapses into the band at 1029 cm^{-1}. Furthermore, the C–H stretching band was shifted to 2919 cm^{1} and the peak due to the CH_2 group was observed around 1375 cm^{-1}.

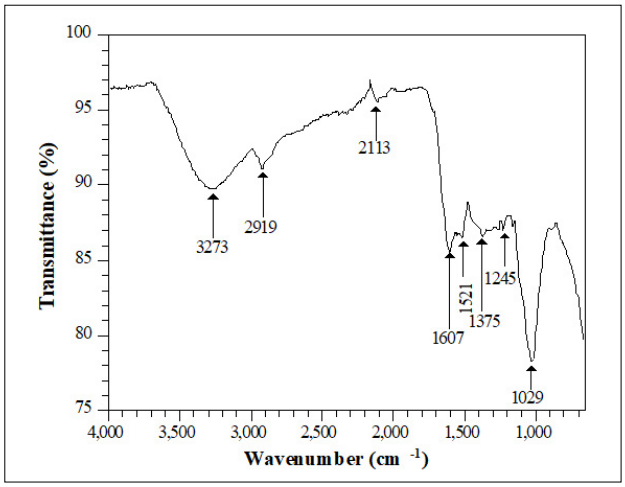


Figure 7: Fourier transform infrared spectrum of SiPCL/TCH confirming the presence of TCH in the SiPCL composite matrix.

The impregnation of the drug into the polymer matrix was responsible for these manifestations. The absence of some FTIR peaks also explains the drop in the baseline intensity of the XRD peaks reported previously in the XRD spectrum of SiPCL/TCH (Figure 5) and, as a result, SiPCL/TCH exhibited lower mechanical properties.

The increase in the intensity of the OH broad band centred at 3273 cm^{-1}, which was attributable to the joint contribution of the OH vibrational stretching of the hydroxyl group in surface water and that in TCH, confirms the presence of TCH in the composite. In addition, due to C=O stretching in the tetracycline rings, a significant peak at 1607 cm^{-1} and a smaller one at 1521 cm^{-1} emerged.[36]

Biodegradability assessment of the SiO_2/PCL-TCH composite

The degradation of SiPCL/TCH after immersion in PBS was observed at 37 °C for a total of 28 days. The relationship in Equation 1 was used to investigate the weight loss and degradability after immersion. Figure 8 shows how the composite degraded over time after immersion in PBS for 28 days. With a loss of about 7.73±0.87% of its starting weight in 24 h, degradation was found to be minor at the start of the investigation, indicating gradual hydrolytic degradation of the sample. On Day 7, the weight reduction climbed to 35.13±1.12%, to 52.18±2.11% on Day 14, 62.33±2.45% on Day 21, and 65.46±1.31% on Day 28.

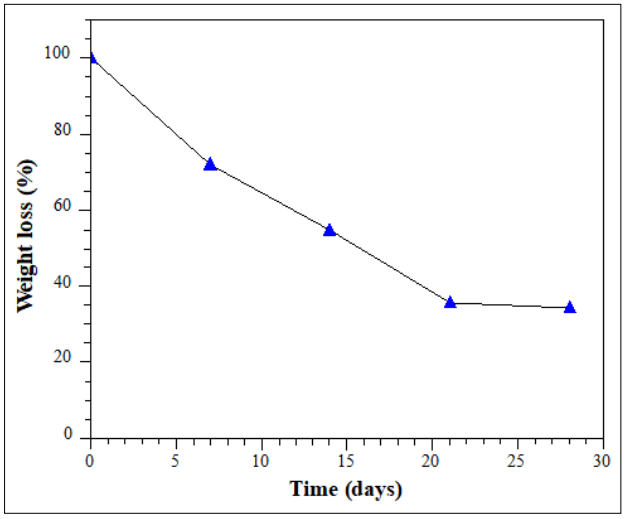


Figure 8: In vitro degradation (weight loss) of SiPCL/TCH soaked in phosphate-buffered saline for 28 days.

After soaking in PBS for 28 days, the degradation result indicates that the majority of the components were lost. PCL is a low-degrading polymer that can be used to construct long-term implanted drug delivery devices.[37] This is frequently determined by the hydrophilic-hydrophobic interactions between the drug and the polymer. The structure and layout of the polymer matrix, pore diameters, and degradation media all influence degradation behaviour. When exposed to water, the PCL, a degradable semi-crystalline synthetic polymer, typically shows surface deterioration. The amorphous portions are targeted first, followed by the crystalline regions for hydrolysis.[38] Random hydrolytic cleavages are also possible in PCL with high molecular weights because they are aliphatic polyesters in nature.[39]

In vitro drug release evaluation

Figure 9 shows the release curve of SiPCL/TCH in PBS at pH 7.4 and 37 °C as assessed by UV–visible spectroscopy at 358 nm. The TCH cumulative release curves showed a biphasic pattern (Figure 9) at the beginning (0–1 day), which is normally associated with a drug incorporated into a polymer as an encapsulating material.[40] At 0–1 h, a release of 9.24±0.93 μg/mL was detected, corresponding to 6.14% of the total concentration of the encapsulated drug. The cumulative release after 3 h was 12.34±1.18 μg/mL, and after 1 day was 65.63±2.13 μg/mL, indicating 12.34% and 43.61%, respectively. After 24 h, the controlled release began and lasted until Day 28. The cumulative releases for 7, 14, and 21 days were 74.10%, 84.43%, and 85.26%, respectively, as shown in Figure 9. The cumulative release peaked at 129.74±4.36 μg/mL after Day 28, accounting for 86.35% of the drug.

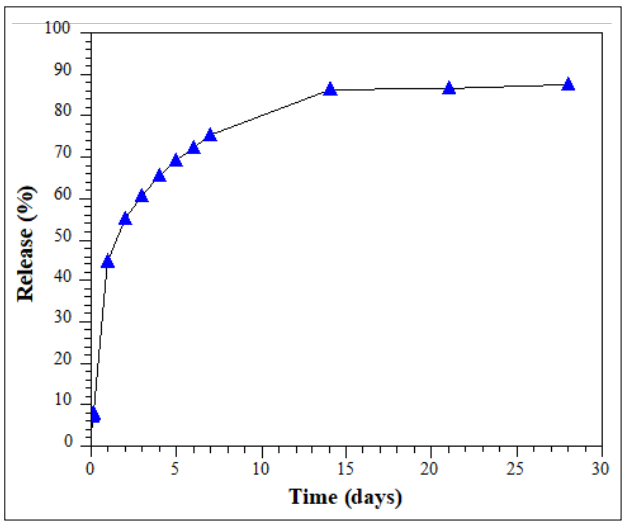


Figure 9: In vitro TCH release profile of SiPCL/TCH immersed in phosphate-buffered saline for 28 days.

On-demand bioavailability of controlled drug release systems gives them a huge advantage over traditional drug administration. The fact that the silica was derived from biomass is significant because, apart from being a cheap source, it can compare to alkoxysilanes as silica precursors.

The silica/PCL composite was made using silica extracted from a bamboo stem at a low cost. The antibiotic tetracycline hydrochloride (TCH) was put into the composite material to act as a drug carrier system for controlled drug release research. The results show that the SiO_2/PCL composite was successfully formed, and the TCH was encapsulated into the composite's matrix. The drug release test yielded a sustained release profile, with the SiO_2 component playing a key role in determining the TCH release kinetics. The rate of degradation of the composite in PBS over 28 days allowed for controlled drug release, which is ideal for a material intended to be used as an in vivo drug release vehicle.

The preparation technique should pose no health or environmental risks when compared to the similar analytical grade silica starting materials based on alkoxysilanes. It is also worth noting that the bamboo plant is abundantly available in Africa and it is underutilised in most regions where it grows, making it an appealing large-scale silica starting material for preparing SiO_2/polymer composites for drug delivery.

Acknowledgements

We thank the Department of Chemical and Food Sciences at Bells University of Technology in Ota, Nigeria, for providing the facilities to undertake this study.

Competing interests

We have no competing interests to declare.

Authors' contributions

E.R.E.: Conceptualisation; student supervision; writing; project leadership; initial draft writing. V.N.A.: Project management; funding acquisition. S.T.S.: Sample analysis; data analysis; validation. O.A.A.: Methodology; data collection.

References

1. Jain KK. Drug delivery systems – An overview. Totowa, NJ: Humana Press; 2008. https://doi.org/10.1007/978-1-59745-210-6_1
2. Wen H, Park K. Oral controlled release formulation design and drug delivery: Theory to practice. Hoboken, NJ: John Wiley & Sons; 2011. https://doi.org/10.1002/9780470640487

3. Nair LS, Laurencin CT. Biodegradable polymers as biomaterials. Prog Polym Sci. 2007;32(8–9):762–798. https://doi.org/10.1016/j.progpolymsci.2007.05.017

4. Dash TK, Konkimalla VB. Poly-ε-caprolactone based formulations for drug delivery and tissue engineering: A review. J Controlled Release. 2012;58(1):15–33. https://doi.org/10.1016/j.jconrel.2011.09.064

5. Wang X, Wang Y, Wei K, Zhao N, Zhang S, Chen J. Drug distribution within poly (E-caprolactone) microspheres and in vitro release. J Mater Process Technol. 2009;209(1):348–354. https://doi.org/10.1016/j.jmatprotec.2008.02.004

6. Schlesinger E, Ciaccio N, Desai TA. Polycaprolactone thin-film drug delivery systems: Empirical and predictive models for device design. Mater Sci Eng C. 2015;57:232–239. https://doi.org/10.1016/j.msec.2015.07.027

7. Pitt CG, Gratzl MM, Jeffcoat AR, Zweidinger RA, Schindler A. Sustained drug delivery systems II: Factors affecting release rates from polycaprolactone and related biodegradable polyesters. J Pharm Sci. 1979;68(12):1534–1538. https://doi.org/10.1002/jps.2600681219

8. Darney PD. Hormonal implants: Contraception for a new century. Am J Obstet Gynecol. 1994;170(5Pt 2):1536–1543. https://doi.org/10.1016/S0002-9378(12)91812-8

9. Chang HI, Williamson MR, Perrie Y, Coombes AG. Precipitation casting of drug-loaded microporous PCL matrices: Incorporation of progesterone by co-dissolution. J Controlled Release. 2005;106(2):263–272. https://doi.org/10.1016/j.jconrel.2005.05.013

10. Chang HI, Perrie Y, Coombes AG. Delivery of the antibiotic gentamicin sulphate from precipitation cast matrices of polycaprolactone. J Controlled Release. 2006;110(2):414–421. https://doi.org/10.1016/j.jconrel.2005.10.028

11. Marchal-Heussler L, Sirbat D, Hoffman M, Maincent P. Poly(epsilon-caprolactone) nanocapsules in carteolol ophthalmic delivery. Pharm Res. 1993;10:386–390. https://doi.org/10.1023/A:1018936205485

12. Chang HI, Wang Y, Perrie Y, Coombe AGA. Microporous polycaprolactone matrices for drug delivery and tissue engineering: The release behaviour of bioactives having extremes of aqueous solubility. J Drug Del Tech. 2010;20(3):207–212. https://doi.org/10.1016/S1773-2247(10)50031-5

13. Sakai-Kato K, Hasegawa T, Takaoka A, Kato, M, Toyo'oka T, Utsunomiya-Tate N, et al. Controlled structure and properties of silicate nanoparticle networks for incorporation of biosystem components. Nanotechnology. 2011;22(20), Art. #205702. https://doi.org/10.1088/0957-4484/22/20/205702

14. Bitar A, Ahmad NM, Fessi H, Elaissari A. Silica-based nanoparticles for biomedical applications. Drug Discov Today. 2012;17(19–20):1147–1154. https://doi.org/10.1016/j.drudis.2012.06.014

15. Kalapathy U, Proctor A, Shultz J. A simple method for production of pure silica from rice hull ash. Bioresour Technol. 2000;73(3):257–262. https://doi.org/10.1016/S0960-8524(99)00127-3

16. Carmona VB, Oliveira RM, Silva WTL, Mattosoa LHC, Marconcini JM. Nanosilica from rice husk: Extraction and characterization. Ind Crop Prod. 2013;43(1):291–296. https://doi.org/10.1016/j.indcrop.2012.06.050

17. Wong DP, Suriyaprabha R, Yuvakumar R, Rajendran V, Chen Y-T, Hwang B-J, et al. Binder-free rice husk-based silicon-graphene composite as energy efficient Li-ion battery anodes. J Mater Chem A. 2014;2(33):13437–13441. https://doi.org/10.1039/C4TA00940A

18. Rangaraj S, Venkatachalam R. A lucrative chemical processing of bamboo leaf biomass to synthesize biocompatible amorphous silica nanoparticles of biomedical importance. Appl Nanosci. 2017;7:145–153. https://doi.org/10.1007/s13204-017-0557-z

19. Musau Z. Bamboo: Africa's untapped potential. The new economic force is generating income, creating jobs and protecting the environment. African Renewal. 2016;April. Available from: https://www.un.org/africarenewal/magazine/april-2016/bamboo-africa%E2%80%99s-untapped-potential

20. Kow KW, Yusoff R, Abdul Aziz AR, Abdullah EC. Physicochemical properties of bamboo leaf aerogels synthesized via different modes of gelation. Appl Surf Sci. 2014;301:161–172. https://doi.org/10.1016/j.apsusc.2014.02.031

21. Sumitha NS, Prakash P, Nair BN, Sailaja GS. Degradation-dependent controlled delivery of doxorubicin by glyoxal cross-linked magnetic and porous chitosan microspheres. ACS Omega. 2021;6(33):21472–21484. https://doi.org/10.1021/acsomega.1c02303

22. Harting R, Johnston K, Petersen S. Correlating in vitro degradation and drug release kinetics of biopolymer-based drug delivery systems. Int J Biobased Plast. 2019;1(1):8–21. https://doi.org/10.1080/24759651.2018.1563358

23. Yu L, Li Y, Zhao K, Tang Y, Cheng Z, Chen J, et al. A novel injectable calcium phosphate cement-bioactive glass composite for bone regeneration. PLoS ONE. 2013;8(4), e62570. https://doi.org/10.1371/journal.pone.0062570

24. Dirna FC, Rahayu I, Maddu A, Darmawan W, Nandika D. Nanosilica synthesis from Betung bamboo sticks and leaves by ultrasonication. Nanotech Sci Appl. 2020;13:131–136. https://doi.org/10.2147/NSA.S282357

25. Wong RSH, Dodou K. Effect of drug loading method and drug physicochemical properties on the material and drug release properties of poly (ethylene oxide) hydrogels for transdermal delivery. Polymers. 2017;9(7):286. https://doi.org/10.3390/polym9070286

26. Gao Y, Li J, Xu C, Hou Z, Yang L. Mechanical properties and drug loading rate of a polycaprolactone 5-fluorouracil controlled drug delivery system. Mater Res Express. 2021;8(9), Art. #095302. https://doi.org/10.1088/2053-1591/ac1fb5

27. Pang Q, Nazar LF. Long-life and high-areal-capacity Li-S batteries enabled by a light-weight polar host with intrinsic polysulfide adsorption. ACS Nano. 2016;10(4):4111–4118. https://doi.org/10.1021/acsnano.5b07347

28. Hao R, Yan Z, Huamin Z, Jinxiang C. Production and evaluation of biodegradable composites based on polyhydroxybutyrate and polylactic acid reinforced with short and long pulp fibres. Cellulose Chem Technol. 2015;49(7–8):641–652. https://www.cellulosechemtechnol.ro/pdf/CCT7-8(2015)/p.641-652.pdf

29. Petersen RS, Nielsen LH, Rindzevicius T, Boisen A, Keller SS. Controlled drug release from biodegradable polymer matrix loaded in microcontainers using hot punching. Pharmaceutics. 2020;12(11):1050. https://doi.org/10.3390/pharmaceutics12111050

30. Blanchard FN. The space group and reference powder diffraction data for tetracycline hydrochloride, $C_{22}H_{24}N_2O_8HCl$. Powder Diffr. 1989;4(2):103–105. https://doi.org/10.1017/S088571560001647X

31. Basu S, Adhiyaman, R. Preparation and characterization of nitrendipine-loaded eudragit RL 100 microspheres prepared by an emulsion-solvent evaporation method. Trop J Pharm Res. 2008;7(3):1033–1041. https://doi.org/10.4314/tjpr.v7i3.14688

32. Phillipson K, Hay J, Jenkins M. Thermal analysis FTIR spectroscopy of poly(-caprolactone). Thermochim Acta. 2014;595:74–82. https://doi.org/10.1016/j.tca.2014.08.027

33. Janarthanan G, Kim IG, Chung E-J, Noh I. Comparative studies on thin polycaprolactone-tricalcium phosphate composite scaffolds and its interaction with mesenchymal stem cells. Biomater Res. 2019;23:1. https://doi.org/10.1186/s40824-018-0153-7

34. Liu X, Ding C, Chu PK. Mechanism of apatite formation on wollastonite coatings in simulated body fluids. Biomaterials. 2004;25(10):1755–1761. https://doi.org/10.1016/j.biomaterials.2003.08.024

35. Adams LA, Essien ER, Adesalu AT, Julius ML. Bioactive glass 45S5 from diatom biosilica. J Sci Adv Mater Devices. 2017;2(4):476–482. https://doi.org/10.1016/j.jsamd.2017.09.002

36. Caminati G, Focardi C, Gabrielli G, Gambinossi F, Mecheri B, Nocentini M, et al. Spectroscopic investigation of tetracycline interaction with phospholipid Langmuir-Blodgett films. Mater Sci Eng C. 2002;22(2):301–305. https://doi.org/10.1016/S0928-4931(02)00217-5

37. Agrawal CM, Ray RB. Biodegradable polymeric scaffolds for musculoskeletal tissue engineering. J Biomed Mater Res. 2001;55(2):141–150. https://doi.org/10.1002/1097-4636(200105)55:2<141::AID-JBM1000>3.0.CO;2-J

38. Bikiaris DN. Nanocomposites of aliphatic polyesters: An overview of the effect of different nanofillers on enzymatic hydrolysis and biodegradation of polyesters. Polym Degrad Stab. 2013;98(9):1908–1928. https://doi.org/10.1016/j.polymdegradstab.2013.05.016

39. Chasin M, Langer R. Biodegradable polymers as drug delivery systems. Drugs Pharm Sci. 1999;92:45.

40. Swornakumari C, Meignanalakshmi S, Legadevi R, Palanisammi A. Preparation of microspheres using poly-3-hydroxybutyrate biopolymer and its characterization. J Environ Biol. 2018;39(3):331–338. https://doi.org/10.22438/jeb/39/3/MRN-315

AUTHORS:
Ahmad Patel[1]
Egerton D.C. Hingston[1]
Molla Demlie[1]

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

CORRESPONDENCE TO:
Egerton Hingston

EMAIL:
hingstone@ukzn.ac.za

DATES:
Received: 18 Oct. 2021
Revised: 04 Mar. 2022
Accepted: 08 Aug. 2022
Published: 30 Nov. 2022

HOW TO CITE:
Patel A, Hingston EDC, Demlie M. Investigation of the suitability of activated and non-activated bentonites from the Imerys Mine (South Africa) for geosynthetic clay liners. S Afr J Sci. 2022;118(11/12), Art. #12566. https://doi.org/10.17159/sajs.2022/12566

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Michael Inggs

KEYWORDS:
bentonite activation, fluid loss, geosynthetic clay liners, smectite clay, swell index

FUNDING:
South African National Research Foundation, University of KwaZulu-Natal

Investigation of the suitability of activated and non-activated bentonites from the Imerys Mine (South Africa) for geosynthetic clay liners

Geosynthetic clay liners (GCLs) have become a suitable substitute for compacted clay liners used for prevention of leachate percolation from landfills and tailings dams into the groundwater system. The characteristics of most bentonites is improved through a process of activation by mixing with soda ash (Na_2CO_3). This paper presents the investigation results of the suitability of bentonites from the Imerys Mine in South Africa for use in GCLs. Both activated and non-activated bentonites were investigated through X-ray diffraction (XRD) analysis and swell index, fluid loss, plate water absorption and Atterberg limits tests. The XRD analysis results indicate that the activated and non-activated bentonite have a smectite content of 58% and 67%, respectively. The swell index of non-activated bentonite was significantly lower than that of activated bentonite. The activated bentonite samples tested at different times subsequent to activation revealed that a minimum of 4 weeks of activation time is required to fully activate it to the ideal soda ash to bentonite ratio of 1:16. The fluid loss tests displayed results slightly above the required minimum of 18 mL because of the low swell index of the bentonites tested. The activated and non-activated bentonites have absorption capacities of 133% and 121%, respectively. The plasticity index of the activated bentonite is 101%, 15% higher than the non-activated bentonite. The overall results concluded that the Imerys bentonite is a medium-quality bentonite with borderline index properties that requires at least 4 weeks of beneficiation to achieve complete activation to suitably be used as GCLs.

Significance:

- The significance of this paper is the beneficiation process of bentonite in the geotextile industry.
- This study is relevant to geotechnical engineers, environmental geologists, engineering geologists and other related professionals working with geosynthetic clay liners whereby bentonite is the key ingredient.
- The study provides an optimum ratio of bentonite to sodium carbonate and the required time for beneficiation.
- We further recommend that quality control measures should be implemented to ensure complete activation of bentonite, including the blending process of different qualities of bentonite.

Introduction

Geosynthetic clay liners (GCLs) have gained much attention over the past decade and have become a common and reliable substitute for compacted clay liners.[1,2] The applications of GCLs span from civil engineering to environmental protection such as liners in water impoundment facilities and as composite liners in landfills.[1,3-5] GCLs are thinly manufactured hydraulic barriers comprising a layer of sodium bentonite encased between two geotextiles or geomembrane sheets.[6] However, the fundamental component of GCLs is bentonite.[1,5,7] Bentonite exists as either sodium bentonite or calcium bentonite, depending on the type of montmorillonite and/or the predominant exchangeable ion it contains. Sodium bentonite has higher water retention characteristics and better swelling properties than calcium bentonite and consequently offers better efficacy as a hydraulic barrier.[8] Generally, the typical sodium bentonite used in GCLs will contain 60–85% montmorillonite.[2,9] GCLs are hydrated to a certain degree of saturation through the uptake of water from the subsoil subsequent to its installation. When the bentonite absorbs water under confinement, it swells, thereby reducing the void ratio. Thus, a lower void ratio can be obtained after the saturated condition, which results in a better performance of the GCL in terms of the limited hydraulic conductivity.[9]

The primary function of GCLs is to act as hydraulic barriers in landfills and tailings dams to contain leachates and tailings, respectively.[1] Therefore, GCL investigations are aligned to properties that relate to hydraulic conductivity, such as fluid loss, free swell and absorption. Sodium bentonites have a high swelling capacity compared to bentonites with low amounts of sodium which have a much lower swelling capacity. However, the low swelling capacity of low sodium bentonites can be improved by treatment with sodium carbonate or soda ash (Na_2CO_3). This process increases the concentration of exchangeable sodium ions within the bentonite, effectively producing sodium exchanged bentonite.[10] The process of mixing bentonite with soda ash is called sodium activation and the resulting bentonite is said to be sodium activated or simply activated bentonite. This is primarily responsible for the high swell index and sealing ability of bentonite.[11]

As various users require bentonite of a certain specification, all mined bentonites undergo some processing prior to sale. Problems such as inadequate swell index defeat the main purpose of a GCL and destroy its primary function as a hydraulic barrier. The Imerys Mine in the Western Cape Province of South Africa mines bentonite for the manufacturing of GCLs. However, information related to fluid loss and swell index performance requirements for use as GCLs has not been studied in detail. Additionally, the comparative performance of the non-beneficiated bentonite relative to beneficiated bentonite and the effect of time on the beneficiation process of bentonite from the mine used in the GCLs has not been researched. Thus, this study reports the investigation results of the suitability

of bentonite from the Imerys Mine for use in GCLs, the comparative performance of the non-beneficiated and beneficiated bentonites and the effect of time on the beneficiation process of bentonite from the mine.

Geosynthetic clay liners

GCLs comprise a thin layer of either sodium bentonite or calcium bentonite, bonded between a geomembrane or geotextile.[1,5,12,13] Those GCLs using geotextiles sandwich the bentonite by needle-punching, stitching or by using a non-polluting adhesive.[1] Needle-punched GCLs comprise an encasing nonwoven filament geotextile where the needle punching process pierces fibres from the upper geotextile to the bottom geotextile.[5] This process entangles the fibres to the bottom geotextile, bonding the sheathing layers together. Bonding may also be achieved through heating, causing the piercing geotextile to fuse to the bottom geotextile. The stitching method involves sewing the geotextiles together using stitching bonded yarns.[5] Bentonite in the geomembrane supported GCL is bonded to the geomembrane using a non-polluting adhesive. The adhesive is mixed with the bentonite and pasted onto a geomembrane.[1,5]

Although there are many types of GCLs, the fundamental difference is the type of bentonite used. The bentonites used in GCLs can either be granular or powdered and sodium or calcium. The main advantages of GCLs are limited thickness, good endurance to differential settlements of underlying soil or waste, simple installation and low cost.[1]

Bentonite deposits at Heidelberg and the Imerys Bentonite Mine

The Imerys Bentonite Mine is located near the town of Heidelberg approximately 230 km east of Cape Town in the Western Cape Province of South Africa (Figure 1). The Mine is situated within the Heidelberg-Riversdale Basin, which hosts the Uitenhage Group rocks.

Geologically, the Uitenhage Group is composed of clastic sediments presented as alluvial fans, braided rivers, and lacustrine sediments.[14] It comprises eight formations, namely: the Hartenbos Formation, the Buffelskloof Formation, the Robberg Formation, the Sundays River Formation, the Brenton Formation, the Infanta Formation, the Kirkwood Formation and the Enon Formation. The bentonite deposits are hosted in the Kirkwood Formation, which is a volcano-sedimentary succession of intercalated sandstones and lacustrine mudstones, with subordinate conglomerates, interbedded with volcanoclastic deposits.[15-17] During the formation of the Heidelberg-Riversdale basin, alkaline volcanic events occurred, leading to the deposition of volcanic ash in a salty lacustrine environment.[18] This saltwater-volcanic ash interaction led to the formation of sodium bentonite. The glass components of the ashes were chemically altered in this low energy environment and consolidated into distinct clay layers that form the Heidelberg-Riversdale bentonite deposits in the study area.[19]

Alteration of volcanic glass may occur through vapour-phase crystallisation, burial diagenesis, contact metamorphism, hydrothermal activity, and by hydrolysis. As the conversion of volcanic glass to smectites involves movement of elements to and from the volcanic glass, leaching of alkali elements and high Mg^{2+}/H^{+} are required to form smectites during the alteration of volcanic glass. Thus, the loss of alkalis and a high magnesium activity promote the formation of smectite.[20]

There are multiple bentonite horizons mined at Heidelberg. These horizons range between 1 m and 1.7 m in thickness and are mostly overlain by mudstone or siltstones.[17] It is estimated that the bentonite reserve at the Heidelberg region is about one million tons.[21]

The method of mining of bentonite at the Imerys Mine is a shallow open pit method, which does not require extensive engineered benches or specific mining techniques. Because bentonite is soft, drilling and blasting are not required, as in the case of conventional mines; however, construction of an access ramp from the overburden is required to allow trucks and excavators into the pit. The pits are excavated to the bentonite layer, which is between 10 m and 20 m below the surface, and the bentonite is removed using an excavator. The excavator simply cuts through the bentonite with the excavator bucket which is then loaded onto tipper trucks for transportation to the processing plant. The pits are refilled and rehabilitated after all the bentonite is mined. As the bentonite quality often varies among several pits, the bentonite from different pits is interlayered to ensure a uniform quality. Thereafter, slices are taken vertically to create a blend of all qualities, and subsequently processed. This method of mixing bentonites of low quality with higher quality bentonites produces a product that has an average acceptable industry standard. It also controls the amount of soda ash that is required, and so to create a fairly consistent mix of bentonite, a fixed amount of soda ash can be added to every batch mined, which also eliminates the need for constantly changing machinery settings and constant quality testing of the bentonite.

Methods and materials

Samples were collected from stockpiles of activated and non-activated bentonite from the Imerys Bentonite Mine. The conversion process of natural sodium bentonite to activated bentonite follows a procedure

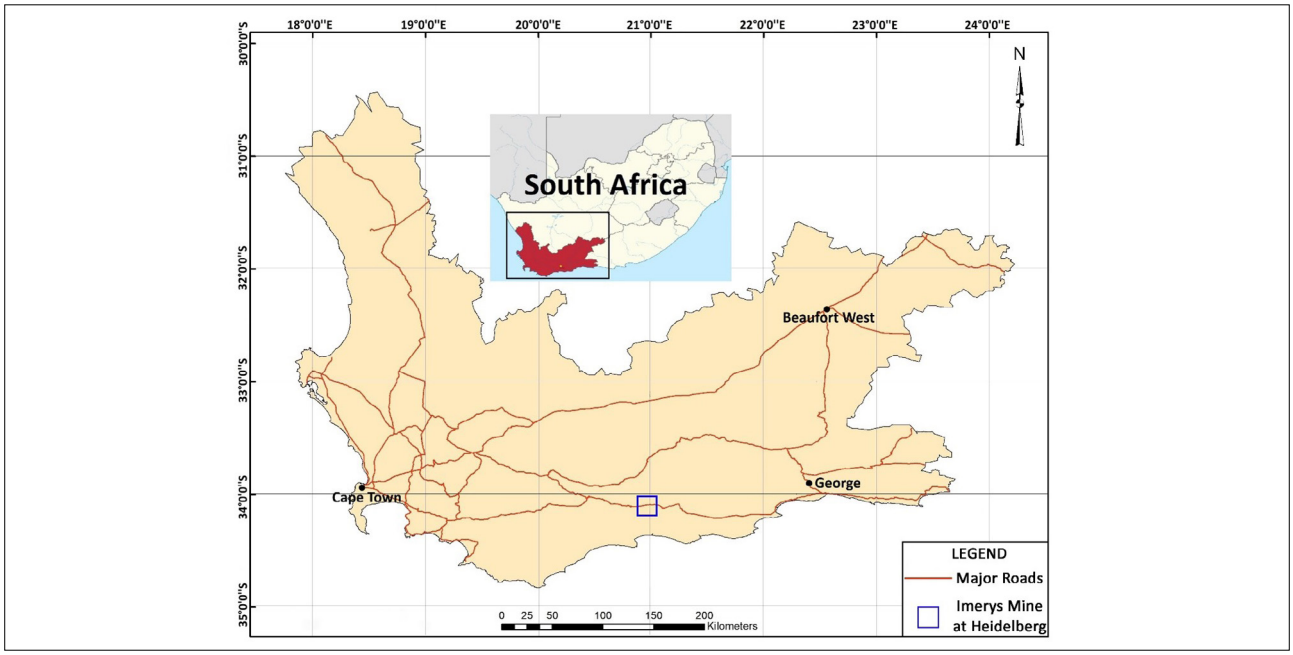


Figure 1: Map of the Western Cape showing the location of the Imerys Mine in Heidelberg.

by which approximately 1–2% of soda ash is added to the bentonite via a feed hopper containing soda ash. The activated bentonite then passes through a rotary drier where all excess moisture is removed and thereafter the dried bentonite is added to a roller crusher where it is milled to a fine powder. Samples of approximately 5 kg from the activated and non-activated bentonite were packed in thick polyethylene bags and then stored in a controlled laboratory environment at 24 °C and 65% relative humidity. Samples were taken randomly from the storage bags for various tests including swell index, fluid loss, plate water absorption, XRD and Atterberg limits following standard procedures.[22-24]

Laboratory analyses

XRD analysis was conducted on the bentonites in order to determine their bulk mineralogical composition. The samples were prepared by drying 10 g of activated and non-activated bentonite in a laboratory oven at 100 °C. The dried bentonite was then crushed with a mortar and pestle and sieved using a 75-μm sieve; 100% passing through the 75-μm sieve were analysed following standard procedures at the XRD Analytical and Consulting Laboratory in Pretoria, South Africa.[25,26]

Swell index test

The swell index test was undertaken following the American Society for Testing and Materials (ASTM) standard.[22] The method involved drying 15 g of bentonite in a laboratory oven at 100 °C, to a constant mass for a minimum of 16 h. The dried sample was then crushed to a fine powder using a mortar and pestle. The powdered bentonite was sieved with the 75-μm aperture sieve and samples passing the 75 μm were used for the tests. A total mass of 2 g of the sieved powdered bentonite was placed in 0.1-g increments every 10 min in a 100-mL graduated cylinder filled with 90 mL of distilled water. Once the entire 2-g sample was added to the cylinder, the water was then topped to 100 mL and left undisturbed for 16 h. The temperature, pH and electrical conductivity (EC) of the water were measured before and after each test. Initially, three samples were tested, with Sample 1 being activated, processed (dried and milled) sodium bentonite; Sample 2 unprocessed, activated sodium bentonite and Sample 3 unprocessed, non-activated sodium bentonite. Activation of Samples 1 and 2 occurred at Imerys Mine, whereby approximately 1% soda ash was added to the bentonite.

To further investigate the effect of time on the beneficiation process, three samples of bentonite were activated in the laboratory with different amounts of soda ash: 2 g, 4 g and 6 g of soda ash was mixed with non-activated bentonite Samples A, B and C, respectively, resulting in ratios of soda ash to bentonite of 1:50, 1:25 and 1:16, respectively. Swell index tests were conducted on each sample within 24 h, and 1 week, 2 weeks and 3 weeks, after activation.

Fluid loss

The fluid loss test evaluates the capability of a bentonite suspension to form a hydraulic barrier. Liu et al.[4] used fluid loss as a quick method to evaluate the hydraulic conductivity of bentonite within GCLs. They reported that the hydraulic performance of GCLs depends directly on the swelling capacity of the bentonite component when in contact with water. Fluid loss tests in this study were conducted at Golder Laboratories in the USA. It was only possible to conduct three fluid loss tests each for activated and non-activated bentonites. The tests were conducted in accordance with the ASTM standard.[23] The fluid loss is calculated as twice the amount of water collected in 22.5 min after the cell has been given 7.5 min to equilibrate, as indicated in Equation 1:

$$\text{Fluid loss (mL)} = 2 \times \text{(mL filtrate volume for the last 22.5 min)} \quad \text{Equation 1}$$

Plate water absorption test

The plate water absorption test was designed to assess the binding ability of bentonite clay binders for iron ore pelletisation.[27] The procedure originated from the ASTM E946-92 testing method, which was initially intended to test the absorption of dried bentonite over a specific period. Balling or pelletising is a process whereby iron ore is processed into a pellet or ball. Bentonite acts as the binding agent in the production of iron ore pellets.[28] The plate water absorption test involves partially submerging a sintered alumina plate in distilled water in a bath. Approximately 2 g of dried and powdered bentonite was placed on a 9-cm-diameter filter paper. The bentonite was spread over the filter paper within a 5-cm-diameter template. The paper and bentonite were placed on the semi-submerged sintered plate to absorb water for 18 h, where the height of the sample above the water surface was 1.2 cm, and the bath was sealed. The water temperature was recorded before the test and 18 h after the test. Then, the bentonite and filter paper were weighed, and the average weight of the wet filter paper was subtracted from the weight of the hydrated filter paper plus the bentonite. This average weight was determined prior to testing by allowing four filter papers to absorb water over an 18-h period without any bentonite. The water absorbed by the bentonite, or the absorption, is the percentage of water absorbed, calculated as a percentage of the dry mass following Equation 2.

$$Absorption, \% = \frac{Ww - Wd}{Wd} \times 100, \quad \text{Equation 2}$$

where W_w is the weight of the hydrated bentonite and W_d is the weight of the dry bentonite in grams.

Atterberg limits

The Atterberg limit is a measure of the critical moisture content, as a percentage of the dry mass, at which a clay soil changes consistency.[29] In this study, Atterberg limits, i.e. the liquid limit, the plastic limit and the plasticity index, were determined using powdered bentonite following the British Standard.[30]

Results and discussion

Mineralogical composition and swelling potential of Imerys bentonite

The results of the XRD analysis for both the activated bentonite and non-activated bentonite are shown in Figure 2. The XRD patterns of the activated bentonite and non-activated bentonite as shown in Figure 2 indicate peaks at 2Θ values of 25, 30 and 60, which are characteristic of montmorillonite. Typically, bentonites contain around 60% montmorillonite.[2] Accordingly, the bentonites from the Imerys Mine can be regarded as typical bentonite as its average smectite content is around 60%. The other peaks in Figure 2 indicate a substantial amount of impurities in both the activated and non-activated bentonites (Figure 3), where the most significant is quartz. According to Gates and Churchman[31], bentonites with impurities will not have a desirable swell index due to the impact of the non-smectite impurities on the physical and chemical properties. The most detrimental types of impurities are non-swelling impurities which diminish the swelling ability of the bentonites.[31] In summary, the quantity of smectite and the chemical composition of bentonite critically influence permeability and performance as a sealing material.[32]

The results of the swell index tests conducted on the activated and non-activated bentonites are shown in Table 1. The measured temperature, pH and EC of the distilled water before and after each test for both the activated and non-activated bentonite are also shown in Table 1. The average swell index obtained for the activated bentonite is 15.6 mL/2 g compared to 9.8 mL/2 g determined for the non-activated bentonite. The results from the swell index tests show that the activated bentonite has a higher swell index than the non-activated bentonite (Table 1). This is because the activated bentonite has additional exchangeable sodium ions due to the activation process. Because sodium ions have a large hydration radius, they are primarily responsible for the swelling of bentonites[11], and this is why the activated bentonite has a higher swell index than the non-activated bentonite. An increase of about 5–6 mL/2 g swell index was observed between the activated bentonite and the non-activated bentonite, amounting to about a 50–70% increase in swell. Figure 4 illustrates the difference in swell index between the activated bentonite and non-activated bentonite.

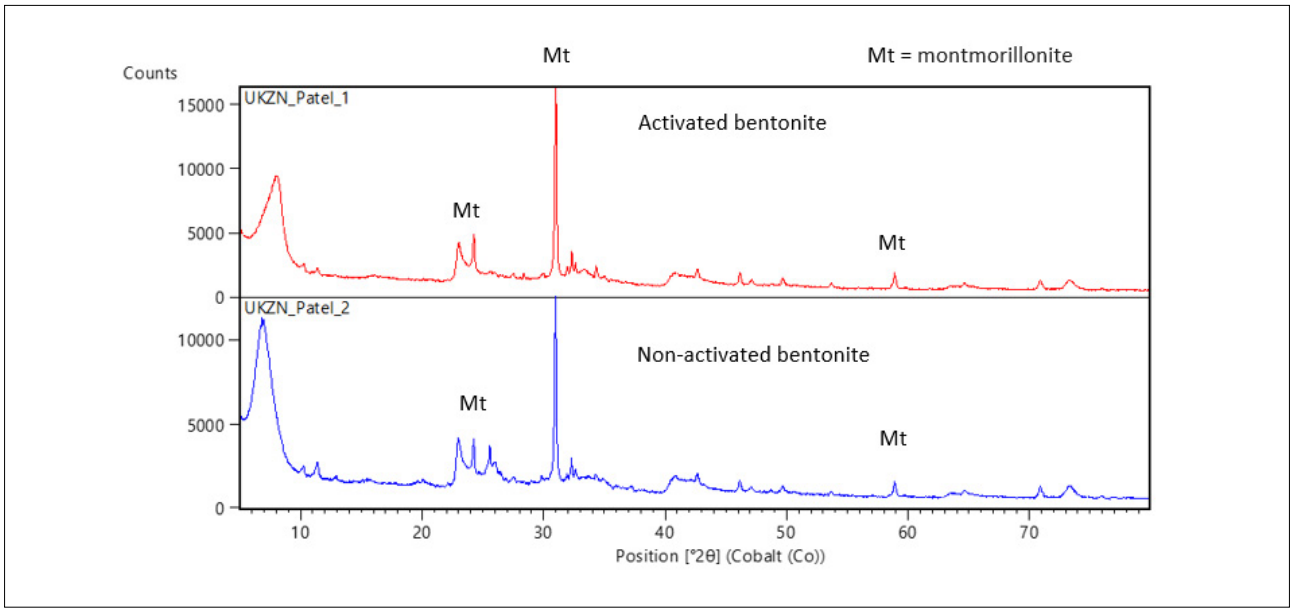


Figure 2: X-ray diffraction pattern of activated bentonite and non-activated bentonite.

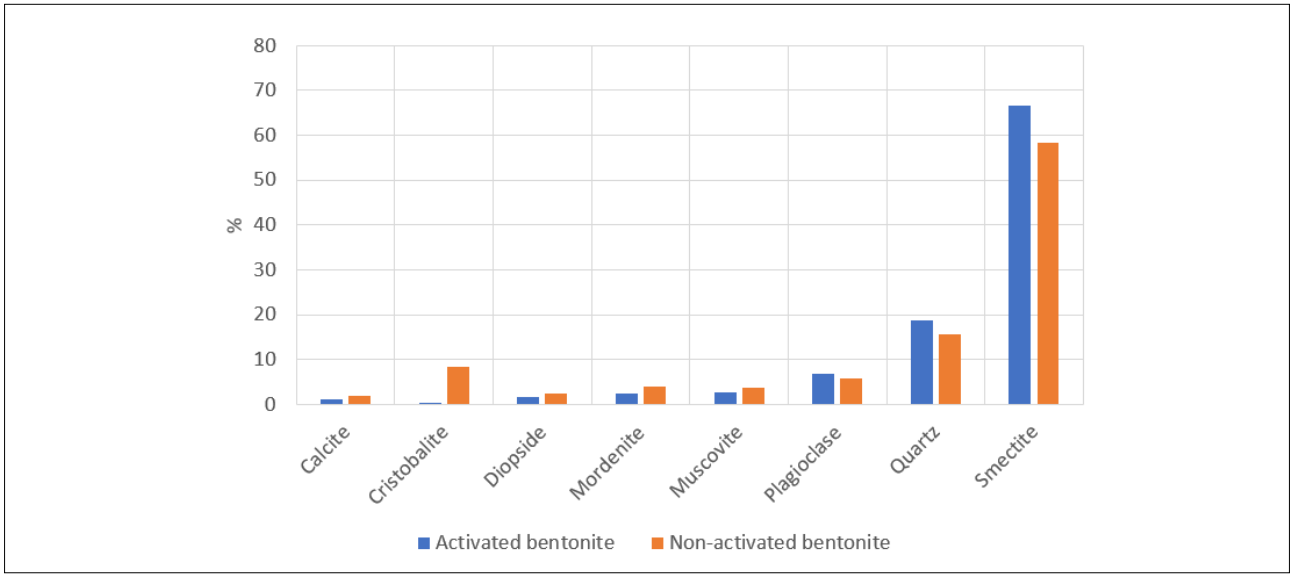


Figure 3: Mineral composition in percentages.

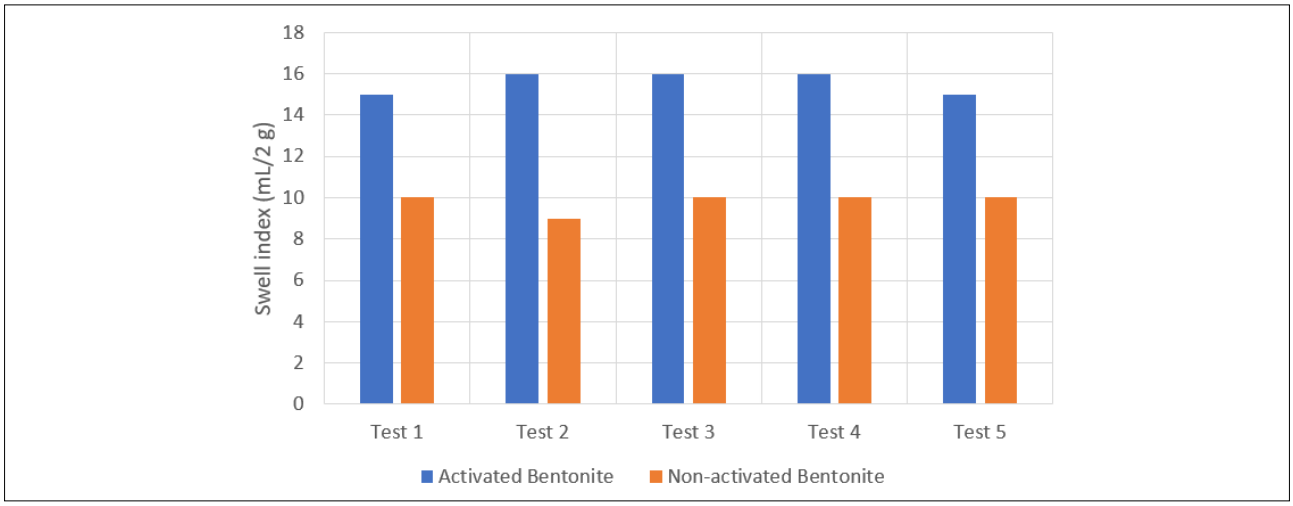


Figure 4: Swell index of activated bentonite and non-activated bentonite.

Table 1: Results from the swell index tests of the activated and non-activated bentonites sampled from the Imerys Mine and measured temperature, electrical conductivity (EC) and pH of the distilled water before and after testing of the activated and non-activated bentonites

	Swell index (mL/2 g)					
	Activated bentonite			Non-activated bentonite		
Test 1	15			10		
Test 2	16			9		
Test 3	16			10		
Test 4	16			10		
Test 5	15			10		
Average	15.6			9.8		
Distilled water	**Temperature (°C)**	**EC (μS/cm)**	**pH**	**Temperature (°C)**	**EC (μS/cm)**	**pH**
Before testing	24	5	9.5	25	3	9.09
After Test 1	23	633	9.82	24.5	134	9.36
After Test 2	23	628	9.9	24.5	155	9.34
After Test 3	23	665	9.91	24.5	165	9.35
After Test 4	23	705	10.1	24.5	194	9.35
After Test 5	23	642	9.9	24	151	9.33

The swell index for the activated bentonite did not reach the minimum requirement of 24 mL/2 g as per the ASTM[22] standard. Further investigation was conducted to ascertain why the swell index did not reach the required specification by undertaking swell index tests of the activated bentonite at specific time intervals after activation. The results indicate that at least 1 month is required after activation before the bentonite acquires the desired swell index. Therefore, as the bentonite activation process is an ionic exchange process whereby the sodium content of the bentonite is enhanced at a molecular level, time is required for the reaction to reach completion.[33] Compared to the activated bentonites at Imerys Mine, the Wyoming bentonite, commonly referred to as MX80 bentonite, has a swell index of approximately 27 mL/2 g.[34] The high swell index of the Wyoming bentonite compared to the activated bentonites tested in this study is attributed to the high montmorillonite content in the Wyoming bentonite.

As seen in Table 2, the temperature and the pH of the solution remained nearly unchanged, indicating that no exothermic or endothermic chemical reaction took place between the de-ionised water and the activated and non-activated bentonites. As distilled water (de-ionised) was used, it can be assumed that no isomorphous ionic substitution took place that could have affected the swelling potential of the bentonite. However, the EC increases substantially during the course of the swell test. The initial ECs of the distilled water used for the activated bentonite and non-activated were 5 μS/cm and 3 μS/cm, respectively. The EC increased to a maximum of 705 μS/cm for the activated bentonite and 194 μS/cm for the non-activated bentonite, indicating that a substantial concentration of ions dissolved into solution. Furthermore, the EC of the activated bentonite solution is higher than that of the non-activated bentonite due to the ionic exchange of sodium allowing the dissolution of accessory minerals into the solution. In this case, the most readily dissolving minerals from natural bentonites are obviously carbonates and sulfates.[35]

Table 2: Swell index of the three samples tested over time after activation

Days after activation	Swell index (mL/2 g)		
	Sample A	Sample B	Sample C
0	10	10	10
1	10	10	22
7	16	19	25
14	18	21	25
21	21	23	26

Effect of time on the beneficiation process

Activation of three samples was conducted in the Laboratory to investigate the effect of time on the beneficiation of bentonites. The results of the swell index tests conducted on the three samples (labelled as A, B and C) at time intervals of 1 day, 7 days, 14 days and 21 days after activation are shown in Table 2. Samples A, B and C had a soda ash to bentonite ratio by mass of 1:50, 1:25 and 1:16, respectively.

Plots of swell index versus time after activation are shown in Figure 5. The plots show that Samples A and B showed no increase in swell index within 24 h, which is explained in terms of the fact that low amounts of sodium activation did not affect the bentonite over such a short period.

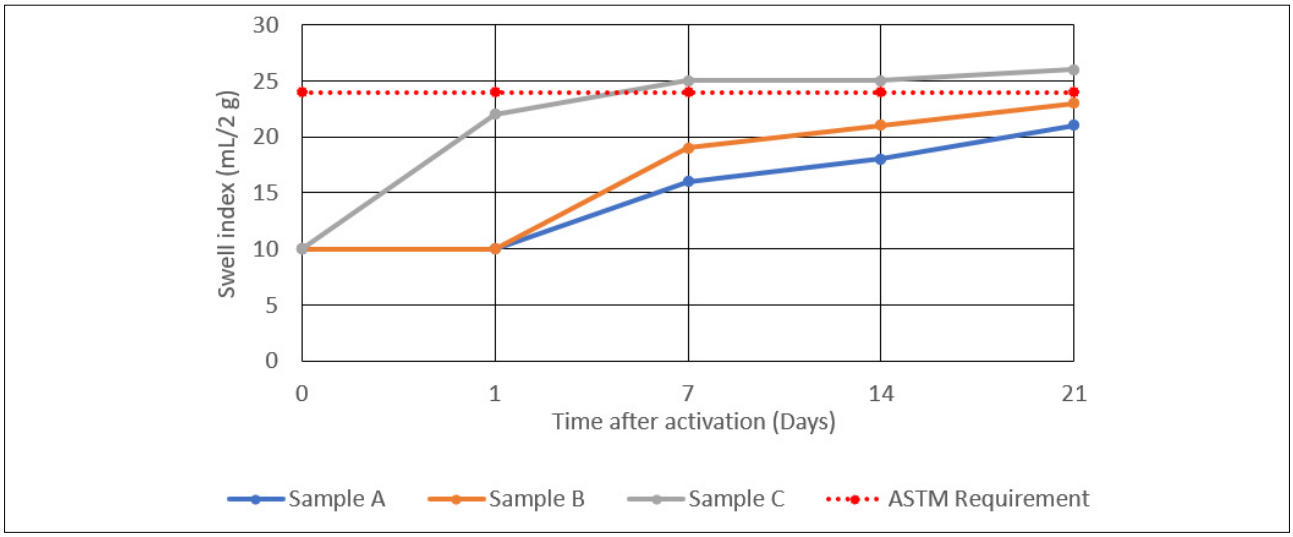


Figure 5: Swell index at various time intervals after activation of samples.

However, Sample C showed an increase in swell index after 24 h of activation, although lower than the required minimum. All three samples showed a marked increase in swell index 7 days after the activation process. Although there was no change in the swell index for Samples A and B after 24 h, Sample B swelled significantly more than Sample A beyond 24 h. The most significant increase in swell for Samples A and B was after 1 week, with Samples A and B increasing by 6 mL/2 g and 9 mL/2 g of swell, respectively, from the initial swell index of 10 mL/2 g. However, 2 weeks of activation time was not sufficient for Samples A and B to reach the required minimum value of 24 mL/2 g. Sample C, however, attained the required minimum value after 1 week of activation. Thus, Sample C represents the best ratio required for activation over a short period, with a soda ash to bentonite ratio of approximately 1:16. Although Sample C achieved the majority of the required swell within the first week after activation, the swell index was enhanced significantly at the start and thereafter levelled off. Samples A and B show a consistent increase in swell index over time but with variable rates. Swell indices for Samples A and B did not reach the benchmark 24 mL/2 g even after 3 weeks of activation.

Fluid loss and plate water absorption test results

The results of the fluid loss tests conducted on the activated and non-activated bentonites are shown in Table 3. Because fluid loss is indicative of the hydraulic conductivity, the activated bentonite is expected to have a lower fluid loss due to a higher swelling and hence sealing ability.[4] The activated and non-activated bentonites tested in this study had average fluid losses of about 23 mL and 27 mL, respectively (Table 3). These measured fluid losses did not meet the 18 mL maximum fluid loss requirement set in the ASTM standard of bentonites used in GCLs.

Table 3: Fluid loss (mL) results of activated and non-activated bentonite

	Average	Range	Number of tests conducted
Activated bentonite	23.2	23.0 – 23.4	3
Non-activated bentonite	26.8	25.7 – 27.9	3

Studies conducted by Shackelford et al.[36] state that an increase in hydraulic conductivity in bentonites is mainly due to limited swelling. Thus, the high fluid losses measured in this study are attributed to a lack of sufficient swell. It is noted that low fluid loss indicates low flow through the bentonite barrier within a given time and therefore implies a low hydraulic conductivity. Furthermore, Lee and Shackelford[37] investigated the effect of the quality of bentonite on its hydraulic conductivity and found that the hydraulic conductivity of high-quality bentonites is three times lower than that of low-quality bentonites. High-quality bentonite is distinguished by its higher montmorillonite content as well as a higher plasticity index.

This study shows that the quality of the raw bentonite produced at the Imerys Mine is of low quality prior to activation and incomplete beneficiation will result in a less desirable product. Therefore, as activation increases the quality of the bentonite to act as a hydraulic barrier, the beneficiation process may require additional time or additional treatment to achieve full activation.

On the other hand, the absorption ability of the bentonite was measured as the water content absorbed by the bentonite after a fixed period. The results of the plate water absorption tests for both the activated bentonite and non-activated bentonite as a percentage of dry mass are presented in Table 4. The average absorption for the activated bentonite was 133.6% whilst for the non-activated bentonite was 121.6%. Both the activated and non-activated bentonites absorbed water well above the dry mass of the respective samples. The activated bentonite absorbed 10% more water than the non-activated bentonite, which is the reason for the relatively increased swell and resultant decrease in the fluid loss, and thereby enhanced its sealing behaviour.[38] Thus, the plate water absorption test indicates the bentonite's ability to swell and it is evident that the addition of soda ash increases the water absorption capabilities of bentonites and consequently its swelling potential and sealing effectiveness.

Table 4: Summary of results from plate water absorption tests

	Water absorption (% of the dry mass)				
	Test 1	Test 2	Test 3	Test 4	Average
Activated bentonite	119.1	129.1	144.1	142.1	133.6
Non-activated bentonite	120.7	117.7	124.1	123.7	121.6

Results of the Atterberg limits tests

The Atterberg limits test results are presented in Table 5. The activated bentonite had a 9% higher liquid limit and 10% higher plasticity index compared to the non-activated bentonite, whilst the plastic limits and the linear shrinkage results were the same (Table 5). The liquid limit of bentonites is controlled by interparticle forces, the dominant being repulsion through osmotic activity which keeps the particles in a fixed configuration and prevents free movement.[39] Strength results from the force of repulsion resisting displacement of particles in the shear plane. On the other hand, as the repulsion is decreased, particles move freely at lower liquid limits. Monovalent ions such as sodium increase the repulsion forces, enabling a higher liquid limit.[39,40] Thus, activation of bentonites increases the concentration of Na^+ thereby increasing the interparticle forces and increasing the swell index. Consequently, the activated bentonite has a higher liquid limit than the non-activated bentonite.

Table 5: Results of Atterberg limits tests of activated and non-activated bentonites

		Atterberg limits (%)	
		Activated bentonite	Non-activated bentonite
Liquid limit	No. of tests	8	8
	Range	130 – 154	128 – 141
	Average	144	135
Plastic limit	No. of tests	3	3
	Range	48 – 51	49 – 53
	Average	50	51
Plasticity index	Average	94	84
Linear shrinkage		25	25

Conclusions

The abundance of bentonite deposits worldwide means GCLs are a common technology in many civil and environmental applications such as landfill hydraulic barriers, environmental protection barriers and as water impoundment liners. Similarly, activated powdered sodium bentonite from the Imerys Mine in the Western Cape Province of South Africa is used in the production of GCLs. The difference in the behaviour of activated and non-activated bentonites and the suitability of the Imerys sodium bentonite for use in GCLs were investigated against the important properties, including the bulk mineralogical composition, swell index, fluid loss, plate water absorption and Atterberg limits. The soda ash to non-activated bentonite ratio and the effect of time on the activation of the bentonite were also investigated.

The results show that the Imerys bentonite is made up of approximately 60% smectite and non-swelling impurities including quartz, calcite, diopside, mordenite, muscovite, and plagioclase. Although the swell index of the activated bentonite was 5 mL/2 g more than that of the non-activated bentonite, both the sodium activated and the non-activated

bentonites failed to reach the required minimum of 24 mL/2 g within the required 24 h. In the case of activated bentonite, this result is attributed to incomplete activation. The fluid loss test for activated and non-activated bentonites produced results higher than the required minimum fluid loss standard of 18 mL, which is attributed to the low quality of the bentonite and its incomplete activation. All the test results indicate that the Imerys bentonite in South Africa is a medium-quality bentonite with borderline index properties that require beneficiation and time to achieve complete activation. The appropriate ratio that suits beneficiation is 1:16 and the required time for beneficiation is at least 4 weeks. Thus, it is recommended that quality control measures should be implemented to ensure complete activation of bentonite, including the blending process of different qualities of bentonite.

Acknowledgements

We thank the Imerys Mine for allowing us to undertake this study, the Kaytech Engineered Fabrics in Pinetown for allowing us to use its laboratory during the course of this study, and the National Research Foundation (South Africa) and University of KwaZulu-Natal for funding the research.

Competing interests

We have no competing interests to declare.

Authors' contributions

A.P.: Data collection, sample analysis, data analysis, validation, writing – initial draft, funding acquisition. E.D.C.H.: Conceptualisation, methodology, validation, student supervision, writing – revisions, project leadership, project management. M.D.: Methodology, validation, writing – revisions, project management.

References

1. Bouazza A. Geosynthetic clay liners. Geotext Geomembr. 2002;20(1):3–17. https://doi.org/10.1016/S0266-1144(01)00025-5
2. Von Maubeuge K. Investigation of bentonite requirements for geosynthetic clay barriers. In: Zanzinger H, Koerner RM, Gartung E, editors. Clay geosynthetic barriers. Lisse: Swets and Zeitlinger; 2002. p. 155–163. https://doi.org/10.1201/9781003078777-19
3. Rowe RK, Orsini C. Effect of GCL and subgrade type on internal erosion in GCLs under high gradients. Geotext Geomembr. 2003;21:1–24. http://dx.doi.org/10.1016/S0266-1144(02)00036-5
4. Liu Y, Gates WP, Bouazza A, Rowe RK. Fluid loss as a quick method to evaluate hydraulic conductivity of geosynthetic clay liners under acidic conditions. Can Geotech J. 2013;51(2):158–163. https://doi.org/10.1139/cgj-2013-0241
5. Kong DJ, Wu H-N, Chai J-C, Arulrajah A. State-of-the-art review of geosynthetic clay liners. Sustainability. 2017;9(11):2110. https://doi.org/10.3390/su9112110
6. Setz MC, Tian K, Benson CH, Bradshaw SL. Effect of ammonium on the hydraulic conductivity of geosynthetic clay liners. Geotext Geomembr. 2017;45:665–673. http://dx.doi.org/10.1016/j.geotexmem.2017.08.008
7. Lin LC, Benson CH. Effect of wet-dry cycling on swelling and hydraulic conductivity of GCLs. J Geotech Geoenviron. 2000;126:40–49. https://doi.org/10.1061/(ASCE)1090-0241(2000)126:1(40)
8. Norrish K. The swelling of montmorillonite. Discuss Faraday Soc. 1954;18:120–134. https://doi.org/10.1039/DF9541800120
9. Seiphoori A, Laloui L, Ferrari A, Hassan M, Khushefati W. Water retention and swelling behaviour of granular bentonites for application in geosynthetic clay liner (GCL) systems. Soils Found. 2016;56(3):449–459. https://doi.org/10.1016/j.sandf.2016.04.011
10. Al-Ani T, Sarapää O. Clay and clay mineralogy. Physical-chemical properties and industrial uses. Geological Survey of Finland Report M19/3232/2008/41. Espoo: Geological Survey of Finland; 2008.
11. Alther G. The qualifications of bentonite as a soil sealant. Eng Geol. 1987;23(3–4):177–191. https://doi.org/10.1016/0013-7952(87)90089-5
12. Sarabian T, Rayhani MT. Hydration of geosynthetic clay liners from clay subsoil under simulated field conditions. J Waste Manag. 2013;33:67–73. http://dx.doi.org/10.1016/j.wasman.2012.08.010
13. Sari K, Chai J. Self healing capacity of geosynthetic clay liners and influencing factors. Geotext Geomembr. 2013;41:64–71. http://dx.doi.org/10.1016/j.geotexmem.2013.08.006
14. Viljoen J, Cawthra H. Lithostratigraphy of the Buffelskloof Formation (Uitenhage Group), South Africa. S Afr J Geol. 2019;122(1):97–104. http://dx.doi.org/10.25131/sajg.122.0009
15. Renne PR, Glen JM, Milner SC, Duncan AR. Age of Etendeka flood volcanism and associated intrusions in southwestern Africa. Geology. 1996;24:659–662. https://doi.org/10.1130/0091-7613(1996)024%3C0659:AOEFVA%3E2.3.CO;2
16. Muir R, Bordy E, Reddering J, Viljoen J. Lithostratigraphy of the Enon Formation (Uitenhage Group), South Africa. S Afr J Geol. 2017;120(2):273–280. https://doi.org/10.25131/gssajg.120.2.273
17. Muir R, Bordy E, Reddering J, Viljoen J. Lithostratigraphy of the Kirkwood Formation (Uitenhage Group), including the Bethelsdorp, Colchester and Swartkops Members, South Africa. S Afr J Geol. 2017;120(2):281–293. https://doi.org/10.25131/gssajg.120.2.281
18. Cole DI, Ngcofe L, Halenyane K. Mineral commodities in the Western Cape Province, South Africa. Cape Town: Council for Geoscience, Western Cape Regional Office; 2014.
19. Christidis GE, Huff WD. Geological aspects and genesis of bentonites. Elements. 2009;5:93–98. https://doi.org/10.2113/gselements.5.2.93
20. Christidis GE. Comparative study of the mobility of major and trace elments during alteration of an andesite and a rhyolite to bentonite, in the islands of Milos and Kimolos, Aegean, Greece. Clays Clay Miner. 1998;46:379–399. http://dx.doi.org/10.1346/CCMN.1998.0460403
21. International Atomic Energy Agency (IAEA). Characterization of swelling clays as components of the engineered barrier system for geological repositories. IAEA-TECDOC-1718. Vienna: IAEA; 2014.
22. American Society for Testing and Materials (ASTM). ASTM D5890: Standard test method for swell index of clay mineral component of geosynthetic clay liners. West Conshohocken, PA: ASTM; 2011. https://doi.org/10.1520/D5890-19
23. American Society for Testing and Materials (ASTM). ASTM D5891: Standard test method for fluid loss of clay mineral component of geosynthetic clay liners. West Conshohocken, PA: ASTM; 2009. https://doi.org/10.1520/D5891_D5891M-19
24. American Society for Testing and Materials (ASTM). ASTM E 946: Standard test method for water absorption of bentonite by the porous plate method. West Conshohocken, PA: ASTM; 1992.
25. Rietveld HM. Line profiles of neutron powder-diffraction peaks for structure refinement. Acta Crystallogr. 1967;22(1):151–152. https://doi.org/10.1107/S0365110X67000234
26. Hanawalt JD, Rinn HW, Frevel LK. Chemical analysis by X-ray diffraction: Classification and use of X-ray diffraction patterns. Powder Diffr. 1986;1(2):2–14. https://doi.org/10.1017/S0885715600011490
27. Mcdonald J, Kawatra SK. Plate water absorption as an assessment tool for organic binders. Miner Process Extr Metall Rev. 2017;38(4):250–253. https://doi.org/10.1080/08827508.2017.1319833
28. Sastry KV, Fuerstenau DW. Mechanisms of agglomerate growth in green pelletization. Powder Technol. 1973;7(2):97–105. https://doi.org/10.1016/0032-5910%2873%2980012-9
29. White WA. Atterberg plastic limits of clay minerals. Am Mineral. 1949;34(7–8):508–512.
30. British Standards Institution. British Standard 1377-2: Methods of test for soils for civil engineering purposes – Part 2. London: British Standards Institution; 1990. https://doi.org/10.3403/00793481
31. Gates WP, Churchman GJ. Bentonites: Their properties, performance and potential. 2006. p. 112.
32. Dananaj I, Frankovská J, Janotka I. The influence of smectite content on microstructure and geotechnical properties of calcium and sodium bentonites. Appl Clay Sci. 2005;28(1–4):223–232. https://doi.org/10.1016/j.clay.2004.02.006

33. Patel A, Hingston EDC, James GM, Mukwevho FV. The effect of time on beneficiation of bentonites used in geosynthetic clay liners. In: Jacobsz SW, editor. Proceedings of the 17th African Regional Conference on Soil Mechanics and Geotechnical Engineering; 2019 October 7–9; Cape Town, South Africa. Cape Town: South African Institute of Civil Engineering: Geotechnical Division; 2019. p. 413–418.

34. Davies CW, Davie CT, Edward CA, White ML. Physicochemical and geotechnical alterations to MX-80 bentonite at the waste canister interface in an engineered barrier system. Geoscience. 2017;7(3):69. http://dx.doi.org/10.3390/geosciences7030069

35. Muurinen A, Lehikoinen J. Porewater chemistry in compacted bentonite. Eng Geol. 1999;54(1–2):207–214. https://doi.org/10.1016/S0013-7952(99)00075-7

36. Shackelford CD, Sevick GW, Eykholt GR. Hydraulic conductivity of geosynthetic clay liners to tailings impoundment solutions. Geotext Geomembr. 2010;28(2):149–162. https://doi.org/10.1016/j.geotexmem.2009.10.005

37. Lee JM, Shackelford CD. Impact of bentonite quality on hydraulic conductivity of geosynthetic clay liners. J Geotech Geoenviron. 2005;131(1):64–77. https://doi.org/10.1061/(ASCE)1090-0241(2005)131:1(64)

38. Norrish K, Quirk J. Crystalline swelling of montmorillonite: Use of electrolytes to control swelling. Nature. 1954;173:255–256. https://doi.org/10.1038/173255a0

39. Warkentin BP. Interpretation of the upper plastic limit of clays. Nature. 1961;190:287–288. https://doi.org/10.1038/190287a0

40. Warkentin BP, Yong RN. Shear strength of montmorillonite and kaolonite related to the interparticle forces. Clays Clay Miner. 1960;9:210–218. https://doi.org/10.1346/CCMN.1960.0090111

AUTHORS:
Andrew N. Green[1,2]
Luke D. Engelbrecht[1]
Kerry Sink[3,4]
Christine F. Mackay[5,6]
Sean T. Fennessy[5,6]
Kathrine Michalsen[7]

AFFILIATIONS:
[1]Geological Sciences, University of KwaZulu-Natal, Durban, South Africa
[2]Environmental Sciences Research Institute, University of Ulster, Coleraine, Northern Ireland, UK
[3]Centre for Biodiversity Conservation, South African National Biodiversity Institute, Cape Town, South Africa
[4]Institute for Coastal and Marine Research, Nelson Mandela University, Gqeberha, South Africa
[5]Oceanographic Research Institute, Durban, South Africa
[6]School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa
[7]Institute of Marine Research (IMR), Bergen, Norway

CORRESPONDENCE TO:
Andrew Green

EMAIL:
greena1@ukzn.ac.za

DATES:
Received: 31 May 2022
Revised: 02 Aug. 2022
Accepted: 02 Aug. 2022
Published: 30 Nov. 2022

HOW TO CITE:
Green AN, Engelbrecht LD, Sink K, Mackay CF, Fennessy ST, Michalsen K. Lifting back the waters: Marine geophysics provides new insights into the uThukela Banks Marine Protected Area. S Afr J Sci. 2022;118(11/12), Art. #14020. https://doi.org/10.17159/sajs.2022/14020

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☒ Not available
☐ Not applicable

EDITORS:
Jennifer Fitchett
Adriaan van der Walt

KEYWORDS:
multibeam bathymetry, seismic reflection, Marine Protected Area, benthic habitats

FUNDING:
NRF African Coelacanth Ecosystem Programme (grant 97968)

Lifting back the waters: Marine geophysics provides new insights into the uThukela Banks Marine Protected Area

Using the first high-resolution geophysical data set collected from the uThukela Banks Marine Protected Area (MPA), we reveal a plethora of hitherto unknown or poorly resolved seabed features. In tandem with several remotely operated vehicle dives, we improve on the previous National Biodiversity Assessment map for the area and reveal a more complex picture of the seabed geology and geomorphology on which the MPA is predicated. The upper slope (-120 m and deeper) is dominated by small canyons, gullies and rills that occasionally extend to the shelf edge and form a series of slumps. Suspected cold-water corals were imaged on the interfluves of the Thukela Canyon. The mid to outer shelf (-60 to -100 m) is mostly rocky, and is composed of Pliocene-age siltstones for the most part. Aeolianite shorelines are found at depths of 60 m and 100 m, in which palaeo-lagoons and parabolic aeolian dune systems are also preserved. These features provide habitat for mesophotic corals and demersal fishes. Overlying and abutting hard rock substrates are unconsolidated sandy sediments that are mobilised by the inshore movement of the Agulhas Current. An inshore mud belt characterised by pockmarks associated with free gas expulsion is mapped for the first time. A well-developed palaeo-drainage pattern is also revealed, posing exciting new opportunities for the study of benthic communities associated with palaeo-estuaries and lagoons now exposed at the seabed. Several new habitats, both inside and out of the MPA boundaries, should form the basis for future research within the MPA, in addition to informing expansions of the MPA.

Significance:

- Using a newly collected geophysical data set, we provide an unprecedented glimpse into the newly proclaimed uThukela Banks Marine Protected Area.
- We reveal a complexity of marine habitats hitherto unknown from previous biodiversity surveys. These habitats include areas of possible expansion given the recognition of keystone species that occur just outside the MPA limits.

Introduction

The 2019 proclamation of 20 new Marine Protected Areas (MPAs) for South Africa heralded a major step forward in the protection of its offshore marine ecosystems, biodiversity and resources. Despite an ambitious increase from 0.4% to 5% in the protection of South African marine and coastal habitats, little is known of the sediment ecology, geology and geomorphology of these areas. Where biological information has limited spatio-temporal resolution, geophysical seafloor attributes may be used as surrogates for ecological processes and species distribution.[1] However, despite the acknowledged strong linkages between offshore geology, geomorphology and benthic species distribution[1], South African MPAs still lack high-resolution geological maps and substrate interpretations. Such knowledge has become increasingly critical for marine spatial planning, including MPA design, implementation, management and expansion in many nations where seabed mapping forms an integral part of such efforts.

The uThukela Banks was identified as a priority area for conservation, based on multiple systematic conservation plans including the Offshore MPA project initiated in 2006.[2] This project used 576 data layers and systematic conservation planning, together with an adaptive stakeholder process to identify priority areas for offshore protection, following which the uThukela MPA was promulgated[3] through a Presidential Oceans Economy Initiative known as Operation Phakisa[4]. The purpose of the uThukela MPA is to protect interconnected coastal, shelf and slope ecosystems, inclusive of sandy, gravelly, reef and submarine canyon habitats.[3] The conservation and protection of ecologically sensitive biodiversity and associated ecological processes is underpinned by vulnerable coral and sponge habitats; threatened species and ecosystems; connectivity pathways; spawning areas; and nurseries. The protected and threatened species include marine mammals (e.g. *Sousa plumbea*), turtles (e.g. *Caretta caretta*), hammerhead sharks (*Sphyrna* spp.), as well as overexploited fisheries species including several species of endemic seabreams (Sparidae), geelbek (*Atractoscion aequidens*), squaretail kob (*Argyrosomus thorpei*), and garrick (*Lichia amia*). The MPA also protects areas of life-history importance for several harvested crustacean species (e.g. *Penaeus indicus*, *Haliporoides triarthrus*) and for migratory fish species.

The complexity of habitat features in any given area is one of the factors that influences both the diversity and the species present, and in turn thus influences the efficiency of no-take MPAs in protecting these species.[5] The area initially identified for protection on the uThukela Banks extended beyond the shelf edge to depths of more than 1000 m, inclusive of the Thukela Canyon, the largest submarine canyon in the area.[6]

In this paper, we present an unprecedented variety of seafloor features comprising unique offshore habitats on the eastern margin of South Africa. These are revealed with a newly acquired geophysical data set, the first from the area since the regional surveys of the 1980s.[7] Using multibeam bathymetric and seismic reflection data, together with direct observational data collected by a remotely operated vehicle (ROV), we reveal the first high-resolution interpretations of the seabed of the uThukela Banks MPA and illustrate advances made in knowledge of the seafloor habitats of the area. We further highlight newly discovered seabed habitats from areas adjacent to the MPA, some of which

may broaden the scope for its future expansion, including areas since removed from the original MPA design. The information we provide will thus feed directly into future iterations of Marine Ecosystems and Mapping for the National Biodiversity Assessment[8] as well as inform national marine spatial planning initiatives[9], and ecosystem-based approaches to fisheries management[10].

Regional setting

The uThukela Banks MPA is situated on the continental shelf of the KwaZulu-Natal (KZN) Bight between Durban and Richards Bay (Figure 1) and spans an area of ~ 4094 km^2. The area has the widest shelf on the east coast of South Africa, with a large fluvial sediment source to landward, and is swept by the vigorous southerly-directed Agulhas Current along its seaward fringes.[11] The benthic habitat is shaped by flows of sediments and nutrients that stem from the coastline and the uThukela River[12-14], while the overlying water column and biota are nourished by the upwelling along the shelf edge and inshore eddy systems[15,16]. Sediment is also supplied to the shelf by the in situ breakdown of several aeolianite palaeo-shorelines that span the MPA.[11] These aeolianites provide the most prominent topographic features of an otherwise flat seabed[7], and they contribute significantly to the area's biological importance and productivity from a habitat and nutrient perspective[17-19].

Abutting and surrounding these aeolianite palaeo-shorelines are sandy sediments that become progressively muddier with proximity to the coastline; these constitute the submerged portions of the wave-dominated Thukela delta.[20] Despite the large quantities of sediment delivered to the shelf, it is comparatively sediment starved when compared to other shelves of the world[12] and much of the mid shelf has been considered to comprise palimpsest gravels, or muds from palaeo-lagoons, exposed by erosion during rising sea levels[11]. These create niche habitats for a variety of benthic organisms not found in comparable quantities elsewhere in South Africa. These organisms include penaeid prawns[21]; fishes favouring turbid, soft-bottom habitats[22]; and infauna that have adapted to fluvial inputs of sediment and detritus, by switching between deposit and suspension feeding depending on the timing of particulate organic matter delivery[23]. Offshore the uThukela River, there are high abundances of hermit crabs (*Paguristes sp.*); these crabs are scavengers that proliferate during high river outflow, making use of the shells of carnivorous *Bullia similis* snails.[24]

Based on the above, the area within the MPA has been classified into several distinct ecosystem types (Figure 2; Table 1).[8] These include a variegated mix of unconsolidated sediments in the mid shelf (KZN Bight and uThukela Shelf Mosaics), a large area of mixed mud and coarse

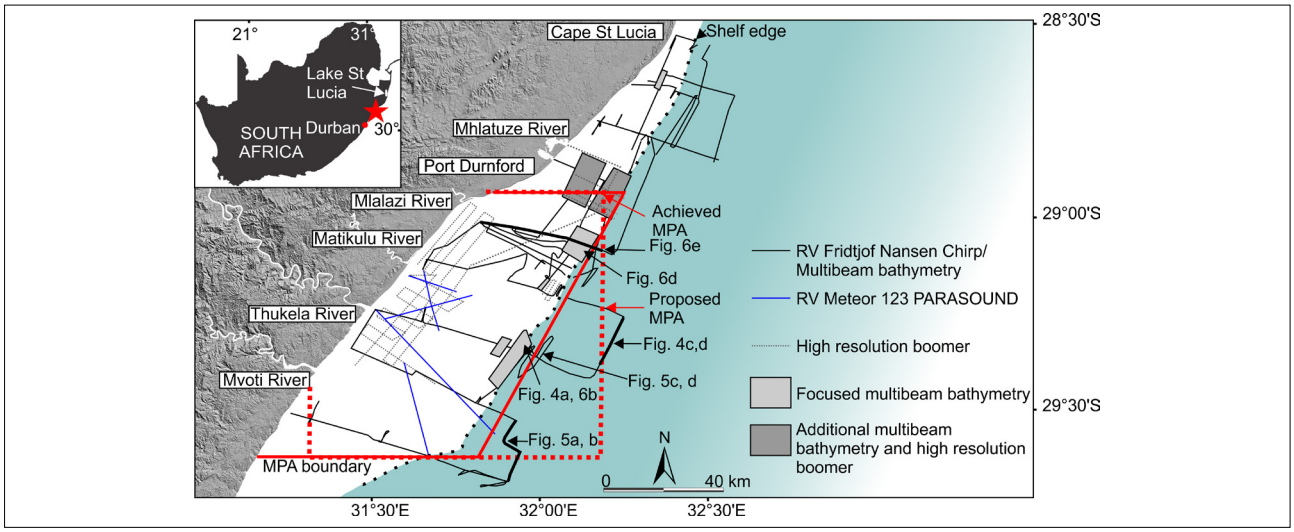


Figure 1: Locality of the uThukela Marine Protected Area (MPA). The solid red line indicates the current MPA limits and the dashed red line indicates the original proposed MPA boundaries. Line tracks for the various geophysical data sets are shown. These include ultra-high resolution Topas seismic reflection and multibeam data (solid black – RV *Fridtjof Nansen*), ultra-high resolution PARASOUND (solid blue) and high-resolution boomer seismic reflection (dotted dark grey). The shelf edge is shown by the black dotted line. Bold lines are seismic reflection profiles shown in figures.

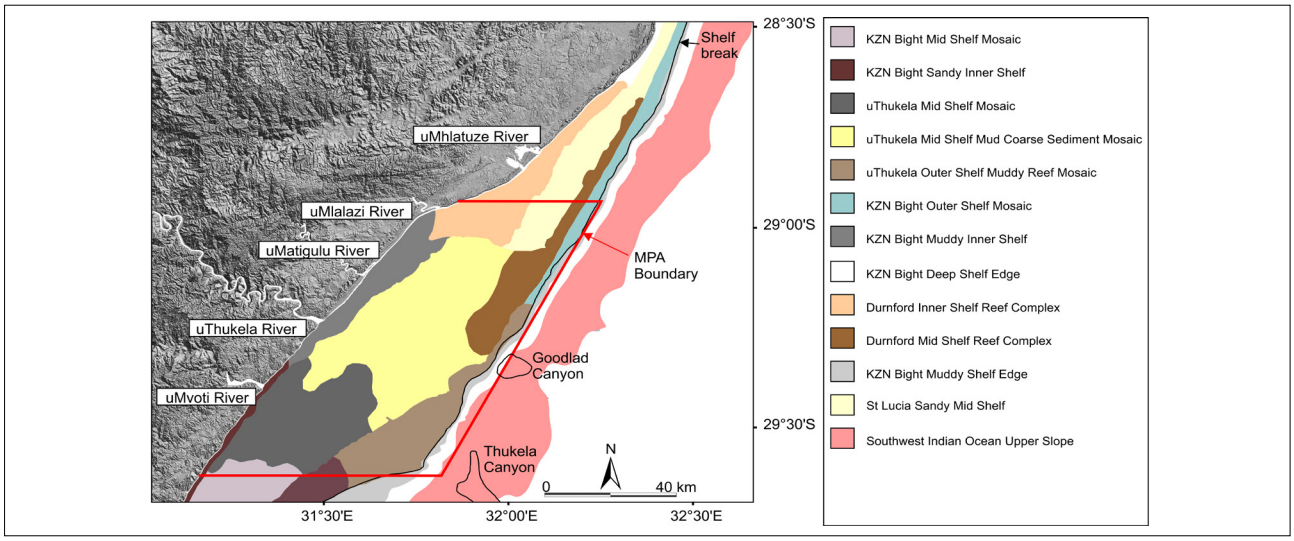


Figure 2: Most recent benthic habitat map for the uThukela Banks Marine Protected Area.[8] Note that the estuaries for each river system are shown.

Table 1: An overview of the seafloor geology/geomorphology and resulting ecosystem types of the uThukela Banks Marine Protected Area

uThukela Marine Protected Area habitat	Description
Durnford Inner Shelf Reef Complex	Durnford reef complex on the inner shelf in the warm, turbid KZN Bight region (-7 to -30 m)
Durnford Mid Shelf Reef Complex	Durnford reef complex in high turbidity mid shelf of the KZN Bight (-30 to -80 m)
KZN Bight Deep Shelf Edge	Unknown seabed type on the shelf edge (-300 to -500 m) in the KZN Bight. Muddy in places and dependent on fluvial inputs
KZN Bight Mid Shelf Mosaic	Mosaic of reef, sand to coarse sand and associated current-influenced pelagic habitat in -30 to -80 m on the mid shelf of the KZN Bight
KZN Bight Muddy Inner Shelf	Muddy seafloor and associated turbid water column on the inner shelf, (back of surf to approximately -30 m) in the KZN Bight
KZN Bight Muddy Shelf Edge	Deep muddy seafloor and associated pelagic habitat on the productive shelf edge of the KZN Bight (-200 to -300 m)
KZN Bight Outer Shelf Mosaic	Mosaic of reef, coarse sand, and associated warm current-influenced pelagic habitat from -80 to -200 m on the outer shelf of the KZN Bight
KZN Bight Sandy Inner Shelf	Sandy seafloor habitat (-7 to -30 m) with associated turbid river-influenced pelagic habitat on the inner shelf of the KZN Bight
Southwest Indian Ocean Upper Slope	Unclassified seabed, -500 to -1000 m in the Southwest Indian Deep Ocean ecoregion
St Lucia Sandy Mid Shelf	Sandy seafloor and productive pelagic habitat in -30 to - 80 m on the mid shelf in the area between Mtunzini and Cape Vidal in the Natal ecoregion; influenced by the St Lucia upwelling cell
uThukela Mid Shelf Mosaic	Mosaic of coarse sand, reef, and associated river-influenced pelagic water column on the mid shelf (-30 to -80 m) in the KZN Bight region
uThukela Mid Shelf Mud Coarse Sediment Mosaic	Mosaic mud and gravel seafloor and associated turbid water column on the mid shelf (-30 to -80 m) of the KZN Bight
uThukela Outer Shelf Muddy Reef Mosaic	Mosaic muddy seafloor and associated turbid water with low profile reef outcrops in -80 to -200 m on the outer shelf of the KZN Bight

sediment (uThukela Mid Shelf Mud Coarse Sediment Mosaic) and an inner muddy shelf (KZN Bight Muddy Inner Shelf). The shelf edge and upper slope are considered muddy (KZN Bight Outer Shelf Mosaic, KZN Bight Muddy Shelf Edge, and uThukela Outer Shelf Muddy Reef Mosaic). Based on a series of seafloor grabs to 200 m depth[11], muds concentrate mostly in the southern portions of the MPA offshore the uMvoti Estuary and give way to coarser sandy and gravelly material in the central and northern areas. The slope is also incised by two canyons, the Thukela and Goodlad Canyons[25], although the locations of their heads have not been accurately constrained.

Two main reef complexes are identified, the Durnford inner shelf and outer shelf reefs, separated from each other by sandy sediments of the St Lucia sandy mid shelf (Figure 2). These form a southward extension of the prominent subaqueous dunes of northern KwaZulu-Natal formed by the Agulhas Current.[26,27]

Methods

Approximately 2000-line km of ultra-high resolution Topas PS18 seismic reflection profiles and multibeam bathymetry were collected between 31 January and 6 February 2018 aboard the RV *Dr. Fridtjof Nansen* (Figure 1). These spanned the southern portions of the uThukela Banks MPA, up to Cape St Lucia and covered an area of 3400 km^2, the purpose of which was to examine key habitat features in the now promulgated MPA, identified by other authors in the planning stages. Seismic reflection data were collected using a Chirp trigger, and were processed using the Kongsberg SBP utility where the final product resolves to $<$ 20 cm in the vertical. Several dedicated blocks of multibeam bathymetry were collected, spanning an area of $\sim$ 180 km^2 of which we present 160 km^2 from the mid to outer shelf and upper slope of the MPA (Figure 1). Multibeam bathymetry was collected using the deep-water Kongsberg EM302 and shallow-water EM710 systems. Sound velocity variations due to salinity and temperature were modelled using a Seabird 911 conductivity-temperature-depth (CTD) sensor. All data were positioned with a differential GPS system of $<$1 m accuracy. Bathymetry data were processed using Beamworx, where spurious soundings were removed and the data reduced to mean sea level using South African Navy tide charts.[28]

These geophysical data were supplemented with ultra-high resolution PARASOUND data and more recently acquired lower resolution seismic reflection data collected between 2018 and 2020.[20] Groundtruthing of the various seabed features on the shelf was performed with a towed camera array aboard the RV *Dr. Fridtjof Nansen*, together with existing ROV footage and grab samples collected under various African Coelacanth Ecosystem Projects (ACEP Natal Bight, ACEP Surrogacy, ACEP Marine Spatial Solutions). Additional multibeam data offshore the port of Richards Bay, collected in 2009 and reported on by other authors[29], were also included. Where rock samples were occasionally retrieved by grabs, these were examined using thin sections and petrographic microscopy. The entire data set was integrated with a lower resolution bathymetry grid of the upper slope[30], to identify key pathways and connections between rivers, canyons and slope gullies in deeper waters.

Results and discussion

Outer shelf and slope

The uThukela MPA outer shelf is characterised by several slumps and mass wasting features (Figure 3 and Figure 4a). The slumps extend upslope to a rocky shelf edge with little to no sediment cover, forming a rugged and variable terrain at water depths of 100–150 m (Figure 4b). Although these are relatively small compared to slumps documented from elsewhere on the South African margin[31,32], they no doubt provide a niche habitat for the area with the exposure of steep and rugged surfaces in the headscarps, and the transport of rocky debris into deeper waters usually dominated by mud. Such varied habitats influence biodiversity, and colonising organisms that themselves create structures that further increase habitat heterogeneity and biodiversity.[33] The slump features may also pose new exploration sites for coelacanth.[34] These features occur between two known living coelacanth habitats at Sodwana Bay[35] and Mzumbe[36], and at a depth similar to habitats reported elsewhere[37].

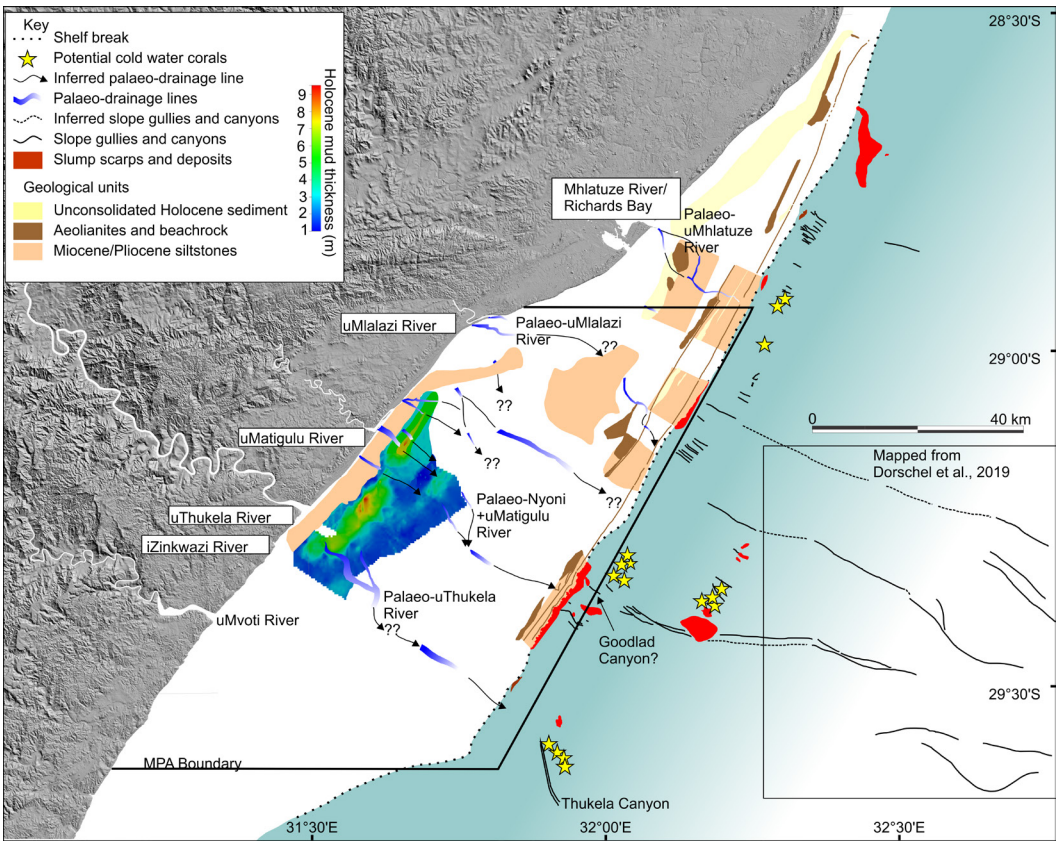


Figure 3: New detailed seafloor geology and geomorphology of the uThukela Banks Marine Protected Area.

Several newly mapped slope-hosted gullies and canyons are also evident from both the multibeam and seismic reflection data (Figure 4a, c and d). These form ≤150-m-wide linear depressions that incise 20–50 m into the seafloor (Figure 4c, d)). The gullies and canyons can be traced further seaward into ~2200 m of water[30] (Figure 3). Where visible, these small canyons and gullies terminate in amphitheatre-shaped depressions along the slumped upper slope.

We identify only one large canyon in the area, the Thukela Canyon (Figure 5a). Our new data resolve the location of the canyon head more precisely than previous low-resolution depictions.[38] It occurs in 755 m of water and gently transitions from the surrounding flat seafloor into a narrow 230-m-wide sinuous depression. The canyon head is significantly narrower than suggested in the National Biodiversity Assessment benthic habitat map[8] and occurs outside of the current MPA limits. Given the importance of protection of submarine canyons as sensitive marine habitats, especially within the initial MPA proposal, this may be a key focal point for further seaward extension of the MPA. The Goodlad Canyon[25,38] is not well resolved. The slope-hosted canyons and slumps shown in Figure 4 match the location reported[38]; however, their scale is much smaller, with a significantly narrower and less entrenched form to that previously thought[25].

Seismic profiles of the mid slope are characterised by 10–50-m-thick, weakly stratified sediment from which several acoustically opaque bodies crop out (Figure 5b). These comprise circular to elongate seafloor features in the bathymetry and occur from -200 m to -800 m, notably straddling the interfluves of the Thukela Canyon on the mid slope (Figure 5b). Although not yet physically observed, they strongly resemble cold-water coral (CWC) accumulations observed by others from similar slope systems worldwide.[39-42] Cold-water corals have a significant role in the zoning of MPAs in South Africa, notably as a keystone species[8], and as long-lived sensitive, habitat-forming species they are global indicators of vulnerable marine ecosystems[42]. These possible new localities may thus prove of critical importance to any plans regarding the further seaward expansion of the uThukela Banks MPA.

Rocky features of the mid to outer shelf

The mid to outer shelf appears to be characterised by mostly rocky outcrops that comprise Miocene- to Pliocene-aged siltstones[43] over which Quaternary-aged aeolianites and beachrocks overlie (Figure 3). The aeolianites and beachrocks comprise a set of submerged shorelines that occur regularly at depths of 60 m, and 100 m (Figure 6a).

-100 m shoreline

The -100 m shoreline is characterised by a single, high relief (≤11 m) and flat-topped ridge of aeolianite that is bordered to landward by a lower elevation seafloor (Figure 6). Directly seaward of the modern uThukela Estuary, and for over 12 km coast-parallel, the landward margin of the -100 m shoreline is backed by semi-circular seafloor depressions ≤2 m deep, separated by cuspate high points on which small 1.5-m-high prograded ridges of aeolianite have developed (Figure 6b). Several small, 230-m-wide and 1-m-deep, unfilled channels enter these depressions and extend to the shelf edge. The depressions have a ≤2-m-thick fill of acoustically opaque material (Figure 6b, section A-B).

Other authors[44] have identified a set of morphologically and stratigraphically similar seafloor features from the Durban Shelf ~83 km to the southwest of this site, which they interpreted as a series of drowned lagoons, separated by cuspate spits in the back barrier on which small prograded barrier sequences have developed. Given the -100 m shoreline's remarkable similarity to these features, and when compared to high-resolution topographic data of the modern coastal waterbodies and dunes of the Maputaland coastline to the north[45], we consider this seafloor feature a drowned lagoon (Figure 6f). This shoreline becomes progressively more linear northward, with small, 150 m-wide crenulations that are similar in appearance to the parabolic dune fields of northern KwaZulu-Natal.[46]

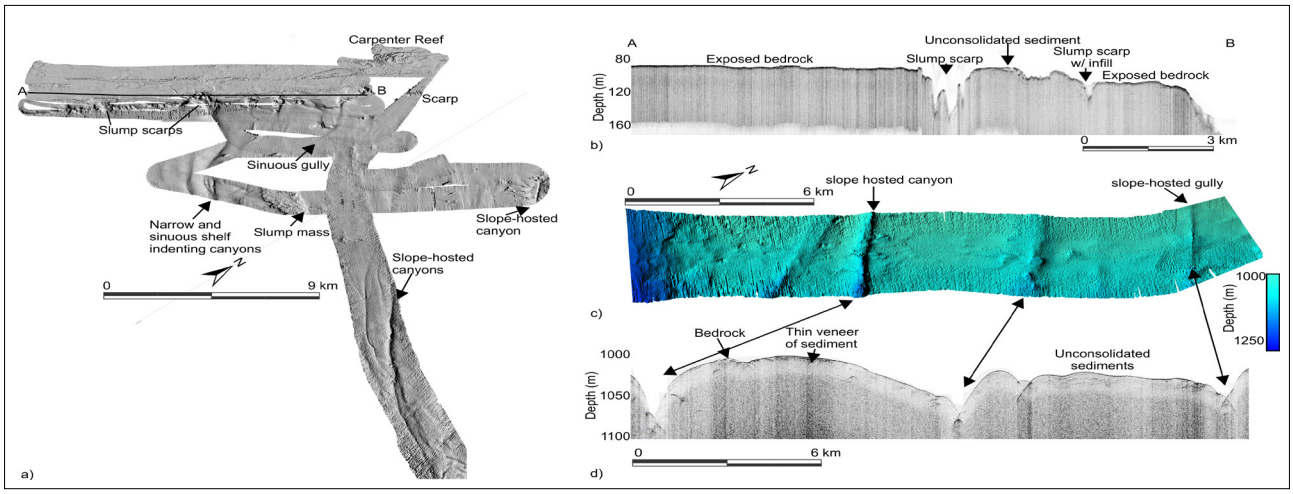


Figure 4: An example of a shelf edge-mass wasting complex. (a) 3D multibeam bathymetry view looking upslope of a series of slumps merging with sinuous gullies and slope-hosted canyons. (b) Seismic reflection profile detailing the hard bedrock of the shelf edge into which the slumps have incised. (c) Multibeam bathymetry showing deeper slope-hosted canyons and gullies. (d) Seismic reflection profile highlighting the interfluves of slope-hosted canyons.

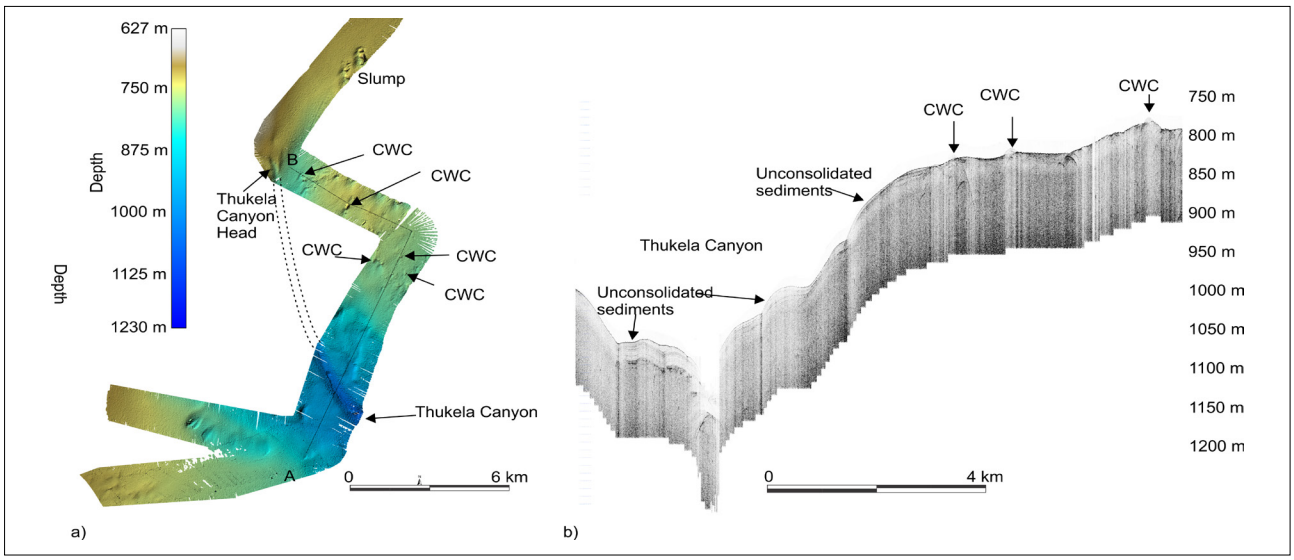


Figure 5: (a) Multibeam bathymetry of the seabed surrounding the Thukela Canyon head with circular seafloor peaks. (b) Seismic reflection profile crossing the Thukela Canyon with suspected cold-water corals (CWCs) cropping out.

-60 m shoreline

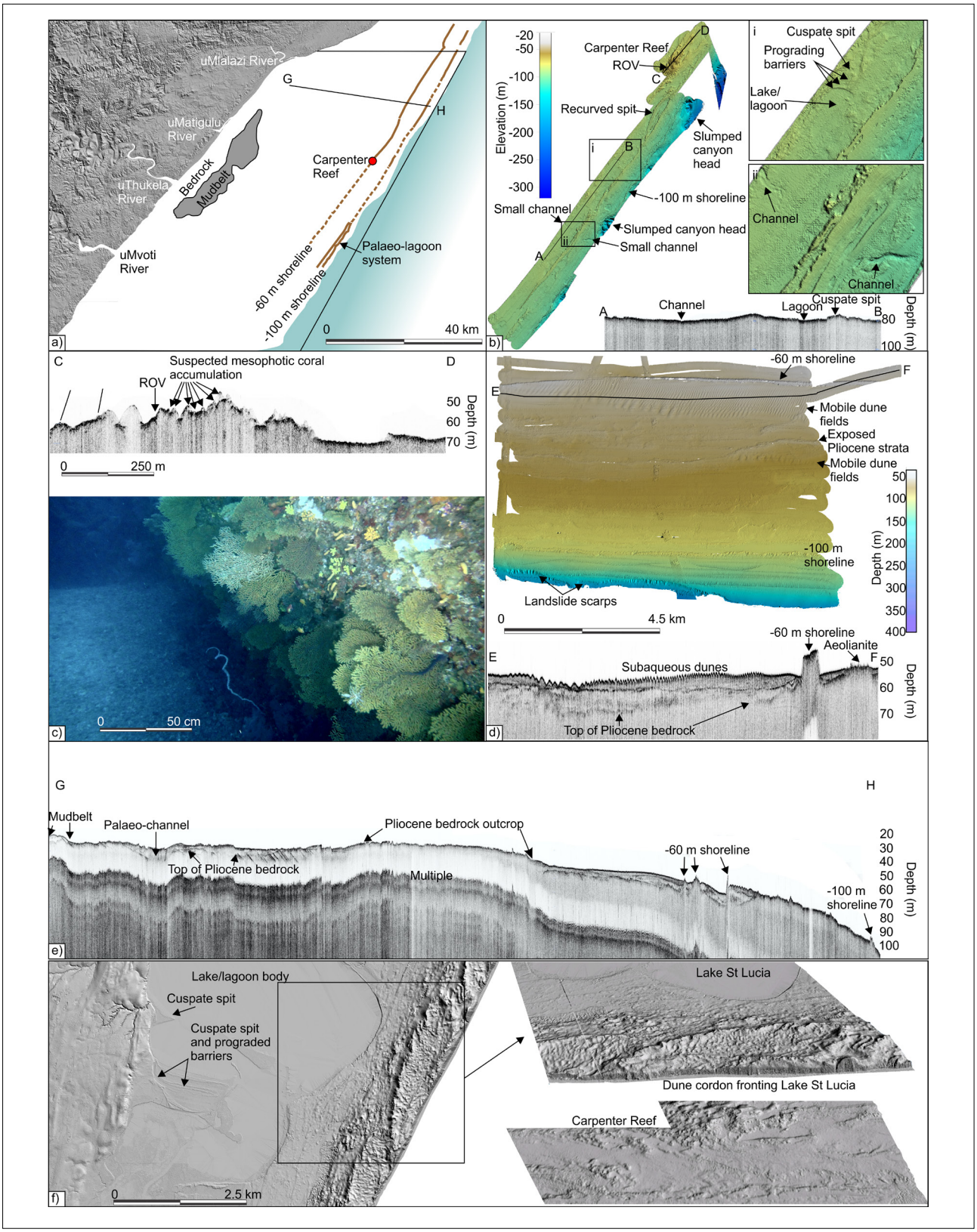
Carpenter Reef is a large aeolianite complex that comprises the more landward part of the -60 m shoreline system rising to -49 m. It is characterised by a relief of between 20 m and 25 m (Figure 6b and c). To date, this is the largest and most prominent seafloor feature mapped on the uThukela Shelf, distinguished by several pinnacles and ridges, the latter of which form parabolic shapes like the contemporary parabolic dunes of the Maputaland coastline (Figure 6f). Comparisons with topographic data from the St Lucia region[45] reveal a remarkable degree of similarity in both onshore and offshore geomorphology, and we consider the seafloor imaged in Figure 6b to be an analogue to the dune systems that border and enclose parts of the coastal waterbodies of the northern KwaZulu-Natal coast (e.g. Lake St Lucia). This shoreline is not well exposed south of the uThukela Estuary and is visible only in the seismic data.

ROV and drop camera transects reveal the -60 m shorelines to have a variety of encrusting mesophotic corals, including dense seafans (probably *Astromuricea fusca*), soft corals *Parasphaerosclera aurea*, dendrophyllid scleractinians[47] and black corals of the genera *Antipathes* and *Stichopathes* (Figure 6c). These are shown in seismic reflection profiles to constitute domed, hyperbolic reflections that form the upper surface, with an acoustically transparent to weakly reflective signature. These mesophotic reefs also host a variety of demersal fishes important in the commercial and recreational linefishery, particularly several species of endemic seabreams (Sparidae), rockcod (Epinephelidae) and kobs (Sciaenidae).[48,49]

Soft sediments of the mid to inner shelf

A ≥40-km-long train of large subaqueous dunes occurs in the mid shelf areas of the MPA, bordered to landward by the -60 m shoreline (Figures 3 and 6d). These constitute the uppermost portions of a ≤10-m-thick, 900-m-wide package of unconsolidated sediment (e.g. Figure 6d, profile E-F). Sediment transport directions are exclusively from north to south, in agreement with the general flow direction of the Agulhas Current, and we consider these as evidence for an occasional, ~5 km in-shelf extension of the Current in the MPA from its usual position along the shelf break.[27]

Most of the inner shelf is dominated by a ≤10-m-thick mud belt that spans the area offshore the uMvoti River to midway between the aMatigulu and uMlalazi Rivers (Figure 3).



ROV, remotely operated vehicle

Figure 6: (a) Shoreline sequences preserved on the shelf of the uThukela Banks Marine Protected Area (MPA). (b) Multibeam bathymetry and seismic reflection profiles of a segmented lagoon complex preserved at -100 m. (c) Seismic reflection profile (top) over Carpenter Reef (location in a) showing raised and acoustically transparent material above the aeolianite surface, confirmed as mesophotic octocoral and black coral assemblages (bottom). (d) Multibeam (top) and seismic reflection profile (bottom) of a linear and narrow shoreline at 60 m water depth, fronted to seaward by unconsolidated sandy sediments shaped into subaqueous dunes. (e) Cross-shelf seismic reflection profile (location in a) detailing the Pliocene outcrop in the mid shelf of the MPA, with the -60 and -100 m shorelines. Note the inshore mud deposits and a deep palaeo-channel in the inner shelf. (f) Comparison between high-resolution terrain models of Lake St Lucia and Carpenter Reef.

Once supporting a shallow-water penaeid prawn fishery[50], it is dominated by smaller sediment-dwelling species (mostly Polychaeta) that have a preference for carbon-rich, muddy environments[24]. Many of the taxa are rare distributions, providing unique biological traits[23] that indicate food resource and habitat diversity[51]. The mud belt occurs between the 15-m and 30-m isobaths, bordered by Pliocene siltstone bedrock outcrop to landward and characterised by high frequency and low to moderate amplitude reflectors that drape the seafloor (Figures 6e and 7a). In the most landward portions of the mud belt, the seismic reflections are blanked (no clear internal stratification) due to the presence of significant quantities of free gas (likely methane) in the sediment.

The mud belt has a variety of gas expulsion pockmarks, evident from the multibeam bathymetry (Figure 7b). They are almost completely circular in shape with a maximum diameter of ~7 m and depths of up to 1 m. The pockmarks do not correspond to any gas chimney or escape features in the seismic data (Figure 7a) and are mostly found where the sediment is gas free or unblanked. This implies a relatively short retention time of the free gas in the sediments, and active gas release to the nearshore environment.

These are the first pockmarks to be recognised from the eastern shelf of South Africa. Despite significant emphasis on the importance of methane seeps from deep-water environments, notably those offshore the Western Cape[52] and especially in the context of niche habitat provision[8], there is nothing known about the habitats provided by these features in South African shallow coastal marine ecosystems. They thus provide an important avenue for future research in the uThukela Banks MPA.

Palaeodrainage

Several palaeo-channels (Figures 3 and 7) can be mapped from the seismic data and are tentatively correlated with the modern drainage patterns of the coast. A prominent palaeo-uThukela River crossed the shelf and appears to have diverted towards the uThukela Canyon head on the mid slope (Figure 3). The palaeo-iNyoni and aMatigulu Rivers merged in the mid shelf and can be traced into the -100 m submerged lagoon where they acted as the feeder valleys to the palaeo-lagoon system.

A few smaller channels are evident between the aMatigulu/iNyoni Estuary and the uMlalazi Estuary; from the existing data it is unclear where they cross the shelf edge. The palaeo-channel of the uMlalazi River is located directly seaward of the modern river and is only evident again in the mid shelf regions. In contrast, the palaeo-uMhlatuze River is very clearly revealed to cross the shelf directly seaward of the Richards Bay port entrance, bifurcating and then re-joining in the mid shelf as a single channel before terminating at the shelf edge. Many of these fluvial pathways were submerged by rising sea levels over the past 18–20 kyr [50], transforming into estuaries now exposed at the seabed due to limited sediment cover. Their exposure likely relates to increased degrees of sediment starvation of the shelf, especially along the outer shelf and at the shelf edge due to current winnowing by the Agulhas Current.[11,53]

Recent work has shown that these palaeo-fluvial courses, and especially exposed relict lagoons and estuaries, act as niche habitats for macrobenthic communities that are unexpected in the shelf environment. For instance, research underway[54] indicates that these systems support remarkable infauna diversity, including endemic species that create biodiversity 'hot spots'. The palaeo-drainage patterns of the MPA thus provide exciting new areas to focus on more detailed examinations of benthic communities in unconsolidated sediment habitats. This is in the context of the exclusion of trawling from the area and the dependencies of various ecosystem functions on contemporary land drainage as a river-dependent marine system.

Summary

The integrated use of marine geophysics has helped reveal several hitherto unrecognised or unknown subtleties in habitat distribution and type in one of South Africa's most recently proclaimed MPAs. When compared to the broad areas of ecozones in the previous habitat map, it is evident that a far greater degree of detail has been derived, with a much higher spatial resolution. In addition, many new features and habitats observed in the area lie outside the current MPA limits.

We show that the incorporation of geological and biological data, together with a range of geophysical approaches, provides invaluable insight for improving South Africa's national map of marine ecosystems. These data will inform MPA management and expansion, marine spatial planning and environmental authorisations that may influence MPA activities in the future. Unconsolidated sandy and muddy shelf ecosystems, palaeo-shorelines, submarine canyons and cold-water corals are often encountered in areas of increasing interest for petroleum and seabed

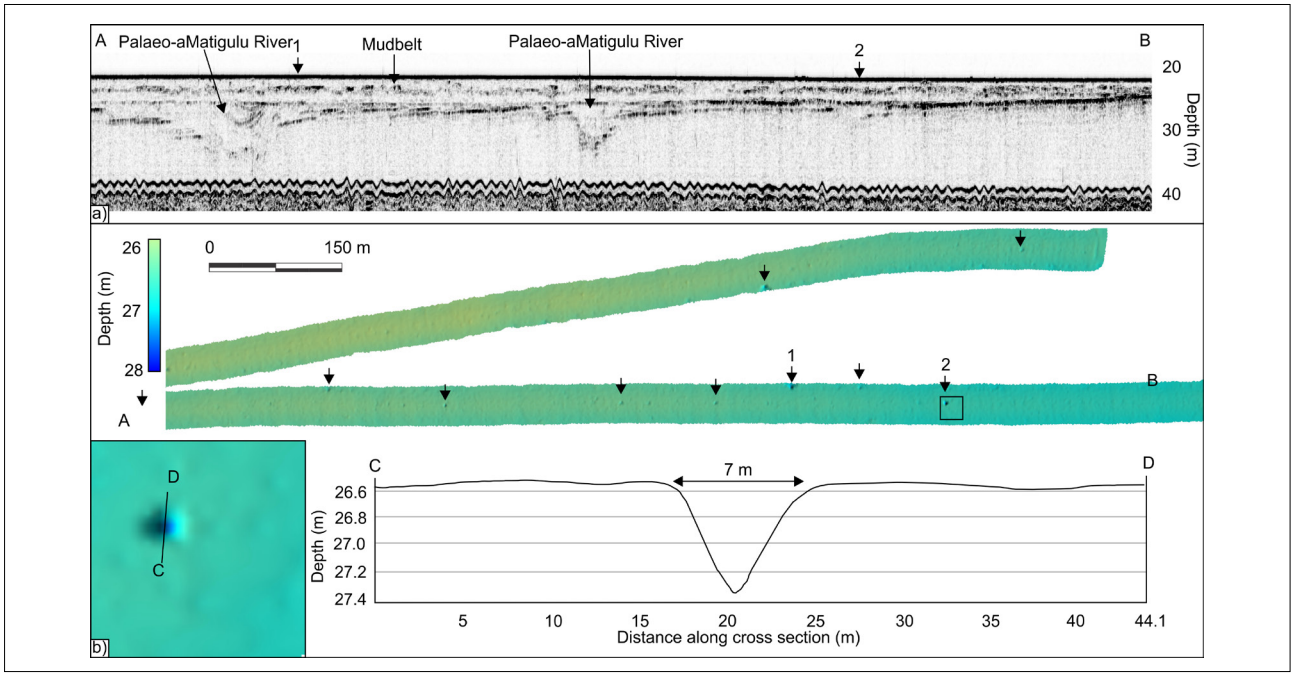


Figure 7: (a) Coast parallel seismic reflection profile offshore the aMatigulu Estuary. Note the occurrence of two prominent river channel meanders which incise by almost 20 m. (b) Corresponding multibeam image of the inner shelf; black arrows highlight pockmarks along the seabed. 1 and 2 are shown in the seismic reflection profile in (a). Cross section C to D highlights a typical pockmark (square and inset).

mining, not just bordering the uThukela Banks MPA, but adjacent to other MPAs throughout the country. We thus urge that the regional mapping campaigns for MPA design and zonation follow a similar approach to ours, where substantial efforts to include geological and geophysical observations are made.

Moving forward, the uThukela Canyon is currently an unprotected ecosystem type with petroleum interests leading to compromises in the early MPA coverage. These resulted in the exclusion of this ecosystem type from the Operation Phakisa MPA network[55]; however, it remains a habitat of importance, especially with regard to the identification of possible cold-water corals. A follow-up benthic sampling campaign that includes ROV or other visual sampling approaches is thus required for confirmation. In addition, further seismic reflection data should be collected in the sandy biomes of the MPA, to better refine our understanding of the palaeo-drainage patterns. Lastly, the use of side-scan sonar to acquire higher-resolution backscatter of the seabed would be helpful in creating automated routines for seabed classification, as has been effectively demonstrated by other authors for the shelves of the Southern and Western Cape.[56,57]

Acknowledgements

For this paper, we used data collected through the scientific surveys with the research vessel *Dr. Fridtjof Nansen* as part of the collaboration between the Food and Agriculture Organization of the United Nations (FAO) on behalf of the EAF-Nansen Programme and South Africa. The EAF-Nansen Programme is a partnership between the FAO, the Norwegian Agency for Development Cooperation (Norad), and the Institute of Marine Research (IMR) in Norway for sustainable management of the fisheries in partner countries and regions. We thank two anonymous reviewers for their inputs, and acknowledge the National Research Foundation (South Africa) and University of KwaZulu-Natal for scholarships for L.D.E.

Competing interests

We have no competing interests to declare.

Authors' contributions

A.N.G.: Conceptualisation, data collection, data analysis, writing initial draft. L.D.E.: data analysis, writing initial draft. K.S.: Writing initial draft. C.F.M.: Data collection, writing initial draft. S.T.F.: Project management, funding acquisition, writing initial draft. K.M.: Project management, funding acquisition.

References

1. Lecours V, Dolan MF, Micallef A, Lucieer VL. A review of marine geomorphometry, the quantitative study of the seafloor. Hydrol Earth Syst Sci. 2016;20(8):3207–3244. https://doi.org/10.5194/hess-20-3207-2016
2. Sink K, Attwood C, Nel D, Lombard AT, Japp D, Branch G, et al. Guidelines for the establishment of an offshore Marine Protected Area network for South Africa. Pretoria: South African National Biodiversity Institute; 2008.
3. Republic of South Africa. Regulations for the management of the iSimangaliso Marine Protected Area in terms of the National Environmental Management: Protected Areas Act (Act No. 57 of 2003). Government Gazette, South Africa R788 (42479); 2019.
4. Sink K. The Marine Protected Areas debate: Implications for the proposed Phakisa Marine Protected Areas Network. S Afr J Sci. 2016;112(9/10); Art. #a0179. https://doi.org/10.17159/sajs.2016/a0179
5. Rees SE, Pittman SJ, Foster N, Langmead O, Griffiths C, Fletcher S, et al. Bridging the divide: Social-ecological coherence in Marine Protected Area network design. Aquat Conserv Mar Freshw Ecosyst. 2018;28(3):754–763. https://doi.org/10.1002/aqc.2885
6. Wiles E, Green AN, Watkeys M, Jokat W, Krocker R. The evolution of the Tugela canyon and submarine fan: A complex interaction between margin erosion and bottom current sweeping, southwest Indian Ocean, South Africa. Mar Pet Geol. 2013;44:60–70. https://doi.org/10.1016/j.marpetgeo.2013.03.012
7. Martin AK, Flemming BW. Physiography, structure add geological evolution of the Natal continental shelf. Lecture notes on coastal and estuarine studies. 1988;26:11–46. https://doi.org/10.1029/LN026p0011
8. Sink KJ, Harris LR, Skowno AL, Livingstone T, Franken M, Porter S, et al. Chapter 3: Marine ecosystem classification and mapping. In: Sink KJ, Van der Bank MG, Majiedt PA, Harris LR, Atkinson LJ, Kirkman SP, et al., editors. South African National Biodiversity Assessment 2018 technical report. Volume 4: Marine realm. Pretoria: South African National Biodiversity Institute; 2019. Available from: http://opus.sanbi.org/bitstream/20.500.12143/6372/3/NBA2018_Vol4_Marine.pdf
9. Department of Forestry, Fisheries and the Environment (DFFE). National data and information report for marine spatial planning: Knowledge baseline for marine spatial planning in South Africa. Cape Town: DFFE; 2021.
10. Jennings S. Indicators to support an ecosystem approach to fisheries. Fish Fish. 2005;6(3):212–232. https://doi.org/10.1111/j.1467-2979.2005.00189.x
11. Green AN, MacKay CF. Unconsolidated sediment distribution patterns in the KwaZulu-Natal Bight, South Africa: The role of wave ravinement in separating relict versus active sediment populations. Afr J Mar Sci. 2016;38(suppl1):S65–S74. https://doi.org/10.2989/1814232X.2016.1145138
12. Flemming B, Hay R. Sediment distribution and dynamics on the Natal continental shelf. Lecture Notes on Coastal and Estuarine Studies. 1988;26:47–80.
13. Bosman C, Ukena R, Leuci R, Smith AM, Sinclair D. Shelf sediments off the Thukela River mouth: Complex interaction between fluvial and oceanographic processes. S Afr J Sci. 2007;103(11):490–492. https://hdl.handle.net/10520/EJC96622
14. De Lecea AM, Fennessy ST, Smit AJ. Processes controlling the benthic food web of a mesotrophic bight (KwaZulu-Natal, South Africa) revealed by stable isotope analysis. Mar Ecol Prog Ser. 2013;484:97–114. https://doi.org/10.3354/meps10311
15. Roberts MJ, Nieuwenhuys C. Observations and mechanisms of upwelling in the northern KwaZulu-Natal Bight, South Africa. Afr J Mar Sci. 2016;38(suppl1):S43–S63. https://doi.org/10.2989/1814232X.2016.1194319
16. De Lecea AM, Cooper R, Smit AJ. Identifying the drivers of the pelagic ecosystem of an oligotrophic bight (KwaZulu-Natal, South Africa) using stable isotopes (δ13C, δ15N) and C: N ratio analysis. Mar Freshw Res. 2015;67(11):1750–1761. https://doi.org/10.1071/MF15256
17. Meyer AA, Lutjeharms JR, De Villiers S. The nutrient characteristics of the Natal Bight, South Africa. J Mar Syst. 2002;35(1–2):11–37. https://doi.org/10.1016/S0924-7963(02)00043-X
18. Hutchings L, Beckley LE, Griffiths MH, Roberts MJ, Sundby S, Van der Lingen C. Spawning on the edge: Spawning grounds and nursery areas around the southern African coastline. Mar Freshw Res. 2002;53(2):307–318. https://doi.org/10.1071/MF01147
19. Samaai T, Sink K, Kirkman S, Atkinson L, Florence W, Kerwath S, et al. The marine animal forests of South Africa: Importance for bioregionalization and marine spatial planning. In: Rossi S, Bramanti L, editors. Perspectives on the marine animal forests of the world. Cham: Springer; 2020. p. 17–61. https://doi.org/10.1007/978-3-030-57054-5_2
20. Engelbrecht L, Green AN, Cooper JA, Hahn A, Zabel M, Mackay CF. Construction and evolution of submerged deltaic bodies on the high energy SE African coastline: The interplay between relative sea level and antecedent controls. Mar Geol. 2020;424, Art. #106170. https://doi.org/10.1016/j.margeo.2020.106170
21. Demetriades NT, Forbes AT. Seasonal changes in the species composition of penaeid prawns on the Tugela Bank, Natal, South Africa. Afr J Mar Sci. 1993;13(1):317–322. https://doi.org/10.2989/025776193784287266
22. Fennessy ST, Villacastin C, Field JG. Distribution and seasonality of ichthyofauna associated with commercial prawn trawl catches on the Tugela Bank of Natal, South Africa. Fish Res. 1994;20(2–3):263–282. https://doi.org/10.1016/0165-7836(94)90087-6
23. Untiedt CB, MacKay CF. Distribution and feeding modes of macrobenthos within three oceanographic feature areas of the KwaZulu-Natal Bight, South Africa. Afr J Mar Sci. 2016;38(suppl1):S91–S104. https://doi.org/10.2989/1814232X.2016.1144651

24. MacKay CF, Untiedt CB, Hein L. Local habitat drivers of macrobenthos in the northern, central and southern KwaZulu-Natal Bight, South Africa. Afr J Mar Sci. 2016;38(suppl1):S105–S121. https://doi.org/10.2989/1814232X.2016.1146631

25. Goodlad SW. Tectonic and sedimentary history of the mid-Natal Valley (SW Indian Ocean) [PhD thesis]. Cape Town: University of Cape Town; 1986. http://hdl.handle.net/11427/23640

26. Flemming BW. Factors controlling shelf sediment dispersal along the southeast African continental margin. Mar Geol. 1981;42(1–4):259–277. https://doi.org/10.1016/0025-3227(81)90166-3

27. Green AN, Flemming BW, Cooper JAG, Wanda TF. Bedform evolution and dynamics of a geostrophic current-swept shelf, northern KwaZulu-Natal, South Africa. Geo-Mar Lett. 2022;42(1):1–22. https://doi.org/10.1007/s00367-021-00722-7

28. South African Navy. 2018 South African tide tables. Cape Town: South African Navy Hydrographic Office; 2018.

29. Green AN, Cooper JAG, Salzmann L. Geomorphic and stratigraphic signals of postglacial meltwater pulses on continental shelves. Geology. 2014;42(2):151–154. https://doi.org/10.1130/G35052.1

30. Dorschel B, Jensen L, Arndt JE, Brummer GJ, De Haas H, Fielies A, et al. The Southwest Indian Ocean Bathymetric Compilation (swIOBC). Geochem Geophys Geosystems. 2018;19(3):968–976. https://doi.org/10.1002/2017GC007274

31. Dingle RV, Robson S. Slumps, canyons and related features on the continental margin off East London, SE Africa (SW Indian Ocean). Mar Geol. 1985;67(1–2):37–54. https://doi.org/10.1016/0025-3227(85)90147-1

32. Green AN, Uken R. Submarine landsliding and canyon evolution on the northern KwaZulu-Natal continental shelf, South Africa, SW Indian Ocean. Mar Geol. 2008;254(3–4):152–70. https://doi.org/10.1016/j.margeo.2008.06.001

33. Buhl-Mortensen L, Vanreusel A, Gooday AJ, Levin LA, Priede IG, Buhl-Mortensen P, et al. Biological structures as a source of habitat heterogeneity and biodiversity on the deep ocean margins. Mar Ecol. 2010;31(1):21–50. https://doi.org/10.1111/j.1439-0485.2010.00359.x

34. Green AN, Uken R, Ramsay P, Leuci R, Perritt S. Potential sites for suitable coelacanth habitat using bathymetric data from the western Indian Ocean. S Afr J Sci. 2009;105(3):151–154. https://doi.org/10.4102/sajs.v105i3/4.68

35. Hissmann K, Fricke H, Schauer J, Ribbink AJ, Roberts M, Sink K, et al. The South African coelacanths – an account of what is known after three submersible expeditions. S Afr J Sci. 2006;102(9):491–500.

36. Fraser MD, Henderson BA, Carstens PB, Fraser AD, Henderson BS, Dukes MD, et al. Live coelacanth discovered off the KwaZulu-Natal south coast. S Afr J Sci. 2020;116(3–4), Art. #7806. https://doi.org/10.17159/sajs.2020/7806

37. Fricke H, Hissmann K, Schauer J, Reinicke O, Kasang L, Plante R. Habitat and population size of the coelacanth *Latimeria chalumnae* at Grand Comoro. In: The biology of *Latimeria chalumnae* and evolution of coelacanths. Dordrecht: Springer; 1991. p. 287–300. https://doi.org/10.1007/BF00007462

38. Dingle RV, Goodlad SW, Martin AK. Bathymetry and stratigraphy of the northern Natal Valley (SW Indian Ocean): A preliminary account. Mar Geol. 1978;28(1–2):89–106. https://doi.org/10.1016/0025-3227(78)90099-3

39. De Mol L, Van Rooij D, Pirlet H, Greinert J, Frank N, Quemmerais F, et al. Cold-water coral habitats in the Penmarc'h and Guilvinec Canyons (Bay of Biscay): Deep-water versus shallow-water settings. Mar Geol. 2011;282(1–2):40–52. https://doi.org/10.1016/j.margeo.2010.04.011

40. Hebbeln D, Van Rooij D, Wienberg C. Good neighbours shaped by vigorous currents: Cold-water coral mounds and contourites in the North Atlantic. Mar Geol. 2016;378:171–185. https://doi.org/10.1016/j.margeo.2016.01.014

41. Hebbeln D, Portilho-Ramos RD, Wienberg C, Titschack J. The fate of cold-water corals in a changing world: A geological perspective. Front Mar Sci. 2019;6, Art. #119. https://doi.org/10.3389/fmars.2019.00119

42. Sundahl H, Buhl-Mortensen P, Buhl-Mortensen L. Distribution and suitable habitat of the cold-water corals *Lophelia pertusa*, *Paragorgia arborea*, and *Primnoa resedaeformis* on the Norwegian continental shelf. Front Mar Sci. 2020;7, Art. #213. https://doi.org/10.3389/fmars.2020.00213

43. Hicks N, Green AN. Sedimentology and depositional architecture of a submarine delta-fan complex in the Durban Basin, South Africa. Mar Pet Geol. 2016;78:390–404. https://doi.org/10.1016/j.marpetgeo.2016.09.032

44. Green AN, Cooper JA, Leuci R, Thackeray Z. Formation and preservation of an overstepped segmented lagoon complex on a high-energy continental shelf. Sedimentology. 2013;60(7):1755–1768. https://doi.org/10.1111/sed.12054

45. Green AN, Humphries MS, Cooper JA, Strachan KL, Gomes M, Dladla NN. The Holocene evolution of Lake St Lucia, Africa's largest estuary: Geological implications for contemporary management. Est Coast Shelf Sci. 2022;266, Art. #107745. https://doi.org/10.1016/j.ecss.2022.107745

46. Jackson DW, Cooper JAG, Green AN. A preliminary classification of coastal sand dunes of KwaZulu-Natal. J Coast Res. 2014;SI70:718–722. https://doi.org/10.2112/SI70-121.1

47. Filander ZN, Kitahara MV, Cairns SD, Sink KJ, Lombard AT. Azooxanthellate Scleractinia (Cnidaria, Anthozoa) from South Africa. ZooKeys. 2021;1066:1–198. https://doi.org/10.3897/zookeys.1066.69697

48. Lamberth SJ, Drapeau L, Branch GM. The effects of altered freshwater inflows on catch rates of non-estuarine-dependent fish in a multispecies nearshore linefishery. Est Coast Shelf Sci. 2009;84(4):527–538. https://doi.org/10.1016/j.ecss.2009.07.021

49. Fennessy ST. Incidental capture of elasmobranchs by commercial prawn trawlers on the Tugela Bank, Natal, South Africa. Afr J Mar Sci. 1994;14(1):287–296. https://doi.org/10.2989/025776194784287094

50. Cooper JAG, Green AN, Compton JS. Sea-level change in southern Africa since the Last Glacial Maximum. Quat Sci Rev. 2018;201:303–318. https://doi.org/10.1016/j.quascirev.2018.10.013

51. Ellingsen KE, Hewitt JE, Thrush SF. Rare species, habitat diversity and functional redundancy in marine benthos. J Sea Res. 2007;58(4):291–301. https://doi.org/10.1016/j.seares.2007.10.001

52. Palan K, Green AN, Andrews B, Sink K, Wiles EA. A morphometric analysis of the fluid flow features of the southern Orange Basin, South Africa. Mar Geol. 2020;423, Art. #106145. https://doi.org/10.1016/j.margeo.2020.106145

53. Flemming BW. Underwater sand dunes along the southeast African continental margin-observations and implications. Mar Geol. 1978;26(3–4):177–198. https://doi.org/10.1016/0025-3227(78)90059-2

54. Mackay CF, Maduna SM. Palaeo-estuarine seafloor features create habitats that support biodiversity hotspots. A dual species- and traits-based approach to detail microbenthic community diversity in KZN coastal waters. Paper presented at: South African Marine Science Symposium; 2022 June 20–24; Durban, South Africa.

55. Sink KJ, Sibanda SM, Fielding P, Skowno AL, Franken M, Harris LR, et al. Chapter 8: Ecosystem protection level. In: Sink KJ, Van der Bank MG, Majiedt PA, Harris LR, Atkinson LJ, Kirkman SP, et al, editors. South African National Biodiversity Assessment 2018 technical report. Volume 4: Marine realm. Pretoria: South African National Biodiversity Institute; 2019. http://hdl.handle.net/20.500.12143/6372

56. Pillay T, Cawthra HC, Lombard AT, Sink K. Benthic habitat mapping from a machine learning perspective on the Cape St Francis inner shelf, Eastern Cape, South Africa. Mar Geol. 2021;440, Art. #106595. https://doi.org/10.1016/j.margeo.2021.106595

57. Pillay T, Cawthra HC, Lombard AT. Integration of machine learning using hydroacoustic techniques and sediment sampling to refine substrate description in the Western Cape, South Africa. Mar Geol. 2021;440, Art. #106599. https://doi.org/10.1016/j.margeo.2021.106599

AUTHORS:
Jenny A. Huggett[1,2]
Johan C. Groeneveld[3,4]
Sohana P. Singh[3]
Sandi Willows-Munro[4]
Ashrenee Govender[4]
Riaan Cedras[5]
Shaun H.P. Deyzel[6,7]

AFFILIATIONS:
[1]Oceans and Coastal Research, Department of Forestry, Fisheries and the Environment (DFFE), Cape Town, South Africa
[2]Department of Biological Sciences, University of Cape Town, Cape Town, South Africa
[3]Oceanographic Research Institute, Durban, South Africa
[4]School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa
[5]Department of Biodiversity and Conservation Biology, University of the Western Cape, Cape Town, South Africa
[6]South African Environmental Observation Network (SAEON), Elwandle Coastal Node, Gqeberha, South Africa
[7]Institute for Coastal and Marine Research, Nelson Mandela University, Gqeberha, South Africa

CORRESPONDENCE TO:
Jenny Huggett

EMAIL:
jenny.huggett@gmail.com

DATES:
Received: 16 Dec. 2021
Revised: 08 July 2022
Accepted: 30 Aug. 2022
Published: 30 Nov. 2022

HOW TO CITE:
Huggett JA, Groeneveld JC, Singh SP, Willows-Munro S, Govender A, Cedras R, et al. Metabarcoding of zooplankton to derive indicators of pelagic ecosystem status. S Afr J Sci. 2022;118(11/12), Art. #12977. https://doi.org/10.17159/sajs.2022/12977

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ Not applicable
☐ Not available
☐ Not applicable

EDITOR:
Sydney Moyo

KEYWORDS:
barcoding reference databases, climate change, essential ocean variables, ocean observations, zooplankton biodiversity

FUNDING:
South African National Research Foundation

Metabarcoding of zooplankton to derive indicators of pelagic ecosystem status

Zooplankton play a key role in marine food webs and carbon cycling and are useful indicators of climate-related changes and ocean health in pelagic ecosystems. Zooplankton are traditionally identified to species through microscopy, but new molecular techniques have enabled the identification of individual specimens (DNA barcoding) or multiple species in the same sample (DNA metabarcoding). Metabarcoding has been tested and refined using zooplankton collected in South African waters for the first time. Challenges to the implementation of DNA-based methods to measure zooplankton biodiversity easily and routinely include an incomplete DNA barcode reference library, logistical complexity and uptake of the new technology by environmental management agencies. These challenges call for a national effort to intensify zooplankton barcoding initiatives and to effectively engage stakeholders in developing a roadmap towards application of DNA-based methods in marine environmental management.

Significance:

- Metabarcoding has been successfully applied to marine zooplankton for the first time in South Africa, demonstrating its potential as a tool to generate ecosystem indicators during routine ocean observations.
- National barcoding efforts must be intensified to provide a comprehensive reference library of zooplankton DNA.
- Effective engagement with stakeholders is required to overcome logistical and policy challenges, and to provide a roadmap towards application of DNA-based methods in marine environmental management.

The need for ocean indicators

Recent warming, acidification and deoxygenation associated with climate change have greatly affected physical and chemical conditions in oceans.[1] Altered marine habitats have led to shifts in the distribution and phenology of organisms, with major implications for biological productivity of marine ecosystems.[1] Yet, human reliance on the goods and services provided by oceans continues to grow.[2] Against a backdrop of climate change and increasing exploitation of marine resources, observations of key ocean indicators are critical to inform policy and support ocean governance and management.[2,3]

The importance of ocean observation systems is recognised by the United Nations Decade of Ocean Science for Sustainable Development (2021–2030), with one of the ten Ocean Decade Challenges being to expand the ocean observing system globally.[4] The Global Ocean Observing System (GOOS) programme has identified a suite of priority physical, biogeochemical and biological ecosystem variables, known as essential ocean variables (EOVs; Supplementary table 1), for routine and sustained observation to assess ocean changes globally, in support of ocean governance.[3] A complementary set of essential biodiversity variables (EBVs)[5], to monitor and reduce biodiversity loss, has been defined by the Group on Earth Observations – Biodiversity Observation Network (Supplementary table 1). Zooplankton biomass and diversity were included as biological EOVs that represent the base of marine food webs.[2] Here, we report on recent molecular methodology (DNA barcoding and metabarcoding) that allows for the rapid and accurate measurement of marine zooplankton biodiversity and relative abundance, and its potential application as a long-term indicator of pelagic ecosystem status in South African coastal and neritic waters. We comment on both methodological and logistical considerations.

Assessing zooplankton biodiversity

Zooplankton play a vital role in the functioning of marine ecosystems. As grazers in pelagic food webs, they provide the main energy pathway from primary producers to higher trophic levels such as fish, squid, and marine mammals. Their excretions fuel the microbial food web and contribute significantly to carbon sequestration via the biological pump.[6,7] Zooplankton are physiologically sensitive to temperature and have short life spans, thus thermal changes are rapidly reflected in their population dynamics.[8] Because they are not fished commercially, changes in zooplankton communities reflect actual environmental or ecosystem-mediated changes, largely unaffected by exploitation trends.[8,9] Zooplankton are therefore excellent 'sentinels' or indicators of change in pelagic ecosystems, with applications extending to climate, fisheries, invasive species, ecosystem health, definition of pelagic ecoregions, biodiversity and ecosystem assessments.[8,9]

Ongoing advancement of DNA barcoding reference data sets (e.g. BOLD, GenBank; Supplementary table 1) coupled with recent metabarcoding technologies now makes rapid and accurate processing of taxonomically complex zooplankton samples logistically feasible. This new technology prompted a shift from traditional microscopy methods to genes as a measure of marine diversity.[10] DNA barcodes can distinguish between visually similar or cryptic species, are independent of life stage, and reduce researcher bias through standardisation of reference systems.[10] DNA barcoding has been applied successfully to global biodiversity studies, including the Census of Marine Zooplankton (CMarZ).[11]

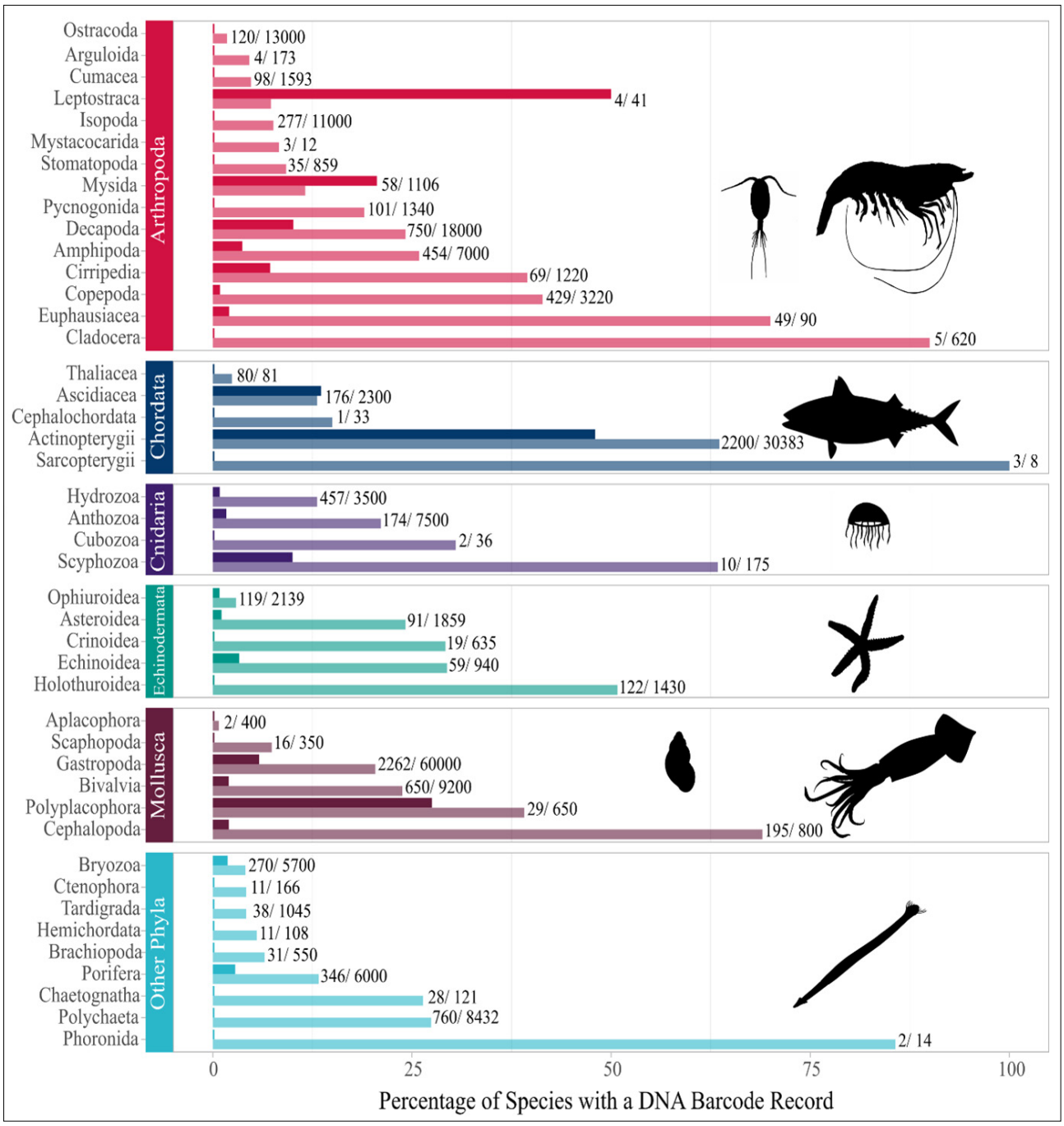
South Africa formally committed to the International Barcode of Life Project (iBOL) in 2011. Barcode reference databases with records of marine zooplankton from South African waters were reviewed recently.[12] Records were proportionally below global levels in nearly all taxa examined and were dominated by species from easily accessible habitats and those with commercially important life phases (i.e. drifting larvae of fish and benthic

invertebrates) (Figure 1). Holoplanktonic species were grossly under-represented, despite making up the bulk of zooplankton biomass and with high importance as ecological indicators. The paucity of holoplanktonic barcodes stemmed from too few specialist taxonomists, the so-called 'taxonomic impediment'.[13]

Metabarcoding uses the same reference databases as barcoding but allows for identification of multiple taxa simultaneously from mixed samples by using high-throughput sequencing platforms.[14] Potential applications of metabarcoding are broad-ranging, including revolutionising biodiversity assessments in any environment for which DNA barcode reference databases are available[15], establishing time-series of diversity[16] and developing biotic indices for routine biomonitoring[17]. The progression from barcoding of individual specimens to metabarcoding of entire communities is now well underway in South Africa, with studies published on diatom communities in the St Lucia estuary[18], bacterial communities in waterholes in the Kruger Park[19] and biomonitoring of freshwater macro-invertebrates[20], among others.

Towards routine biodiversity monitoring

The adoption of molecular approaches such as metabarcoding in marine environmental management requires an iterative 'translational molecular ecology' approach (constant two-way communication between scientists and stakeholders).[17] Recent applications that demonstrate this process internationally include routine monitoring of ichthyoplankton, biosecurity monitoring for non-indigenous species, and ecological status assessments.[17] As part of the translational molecular ecology approach,



Source: Adapted from Singh et al.[12] with permission, ©NISC (Pty) Ltd.

Figure 1: The relative percentages of DNA barcode records available for marine zooplankton taxa, globally (pale bars) and for South Africa (dark bars). Numbers next to the bars are the numbers of species known locally/globally.

the development of and adherence to standardised protocols to guarantee data comparison across spatial and temporal scales are crucial.[17]

Metabarcoding of marine zooplankton in South Africa has been optimised to follow best-practice protocols set by the international Scientific Committee on Oceanic Research working group MetaZooGene (Supplementary table 1). Optimisation challenges were the design of taxon-specific mini-barcode primers to increase species detection rates[21], experimental validation of primer cocktails to test their efficiency in detecting rare species[22], and comparison of species identities obtained from metabarcoding and microscopy[12]. The strategic importance of an expanded DNA barcode reference database for the region is recognised by the Foundational Biodiversity Information Programme of the South African National Research Foundation. This Programme currently funds integrative molecular morphology projects to increase regional reference databases (including for zooplankton), with new records uploaded to open-access databases such as BOLD and GenBank.

Further incorporation of zooplankton biodiversity in ocean observation programmes as biological EOVs faces several logistical challenges in South Africa. Key among these are: access to sea-going vessels and sampling gear, laboratories for sample processing, technical and scientific expertise across various disciplines (genetics, taxonomy, bioinformatics, biodiversity, oceanography) and data management from the point of observation to implementation by users. To evaluate the underlying science, operational costs and benefits, a small-scale pilot project was initiated in 2018, along a single cross-shelf transect at the KwaZulu-Natal Bight Sentinel Site (Supplementary table 1; Supplementary figure 1). The pilot project is a collaboration between several academic institutions, which are in turn funded by the Department of Science and Innovation and the National Research Foundation.[12,21,22] Key findings were that costs, coordination and uptake will be the main challenges for a prospective long-term programme.

In addition to a successful 'proof of concept' exploratory project and an improved barcode reference database, progress towards the routine use of metabarcoding in monitoring of the marine environment in South Africa will require a 'translational molecular ecology' process to facilitate its uptake at environmental management and policy levels.[17] The role of national environmental observation agencies, primarily the South African Environmental Observation Network (SAEON) and the Department of Forestry, Fisheries and the Environment (DFFE), in providing operational budgets and coordination, and as long-term custodians of data and indices, will be critical for actionable progress beyond exploratory research. Our Research Letter describes the initial development of a DNA-based method for biodiversity assessments of South African pelagic ecoregions, including enabling conditions for its uptake in marine environmental management.

Acknowledgements

J.A.H. acknowledges support from the Scientific Committee on Oceanic Research (SCOR) WG157 MetaZooGene. The South African Institute of Aquatic Biodiversity (SAIAB) is thanked for the use of their research vessel (RV) *Phakisa* and sampling equipment, and the assistance of the vessel's crew. The KwaZulu-Natal Research and Innovation Sequencing Platform (KRISP) is thanked for undertaking Illumina sequencing of mixed zooplankton samples. The National Research Foundation (NRF) is thanked for doctoral and post-doctoral financial support for A.G. and S.P.S. The provision of research platforms (Coastal Crafts, Sentinel Site Instrumentation and Data Management) by the Shallow Marine and Coastal Research Infrastructure (SMCRI) of the Department of Science and Innovation (DSI) hosted by NRF-SAEON Elwandle Node and NRF-SAIAB is acknowledged.

Competing interests

We have no competing interests to declare.

Authors' contributions

J.A.H.: Conceptualisation, methodology, writing – the initial draft, writing – revisions, project leadership, project management. J.C.G.: Conceptualisation, methodology, writing – the initial draft, writing – revisions, project leadership, project management. S.P.S.: Methodology, data collection, data analysis, writing – the initial draft. S.W.-M.: Methodology, data analysis, writing – the initial draft, student supervision. A.G.: Methodology, data collection, data analysis, writing – the initial draft. R.C.: Methodology, data collection, data analysis, writing – the initial draft. S.H.P.D.: Methodology, data collection, data analysis, writing – the initial draft.

References

1. IPCC. The ocean and cryosphere in a changing climate: Special report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2019. Available from: https://www.ipcc.ch/srocc/
2. Muller-Karger FE, Miloslavich P, Bax NJ, Simmons S, Costello MJ, Sousa Pinto I, et al. Advancing marine biological observations and data requirements of the complementary essential ocean variables (EOVs) and essential biodiversity variables (EBVs) frameworks. Front Mar Sci. 2018;5:211. https://doi.org/10.3389/fmars.2018.00211
3. Miloslavich P, Bax NJ, Simmons SE, Klein E, Appeltans W, Aburto-Oropeza O, et al. Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. Glob Change Biol. 2018;24:2416–2433. https://doi.org/10.1111/gcb.14108
4. United Nations Decade of Ocean Science for Sustainable Development. The ocean decade [webpage on the Internet]. c2021 [cited 2021 Dec 01]. Available from: https://www.oceandecade.org/
5. Pereira HM, Ferrier S, Walters M, Geller GN, Jongman RHG, Scholes RJ, et al. Essential biodiversity variables. Science. 2013;339:277–278. https://doi.org/10.1126/science.1229931
6. Passow U, Carlson CA. The biological pump in a high CO_2 world. Mar Ecol Prog Ser. 2012;470:249–271. https://doi.org/10.3354/meps09985
7. Steinberg DK, Carlson CA, Bates NR, Goldthwait SA, Madin LP, Michaels AF. Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea. Deep Sea Res I. 2000;47:137–158. https://doi.org/10.1016/S0967-0637(99)00052-7
8. Richardson AJ. In hot water: Zooplankton and climate change. ICES J Mar Sci. 2008;65:279–295. https://doi.org/10.1093/icesjms/fsn028
9. Batten SD, Abu-Alhaija R, Chiba S, Edwards M, Graham G, Jyothibabu R, et al. A global plankton diversity monitoring program. Front Mar Sci. 2019;6, Art. #321. https://doi.org/10.3389/fmars.2019.00321
10. Laakmann S, Blanco-Bercial L, Cornils A. The crossover from microscopy to genes in marine diversity: From species to assemblages in marine pelagic copepods. Phil Trans R Soc B. 2020;375(1814), Art. #20190446. https://doi.org/10.1098/rstb.2019.0446
11. Bucklin A, Peijnenburg KTCA, Kosobokova KN, O'Brien TD, Blanco-Bercial L, Cornils A, et al. Toward a global reference database of COI barcodes for marine zooplankton. Mar Biol. 2021;168, Art. #78. https://doi.org/10.1007/s00227-021-03887-y
12. Singh SP, Groeneveld JC, Huggett J, Naidoo D, Cedras R, Willows-Munro S. Metabarcoding of marine zooplankton in South Africa. Afr J Mar Sci. 2021;43(2):147–159. https://doi.org/10.2989/1814232X.2021.1919759
13. Engel MS, Ceríaco LMP, Daniel GM, Dellapé PM, Löbl I, Marinov M, et al. The taxonomic impediment: A shortage of taxonomists, not the lack of technical approaches. Zool J Linn Soc. 2021;193:381–387. https://doi.org/10.1093/zoolinnean/zlab072
14. Cristescu ME. From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. Trends Ecol Evol. 2014;29:566–571. https://doi.org/10.1016/j.tree.2014.08.001
15. Bucklin A, Lindeque PK, Rodriguez-Ezpeleta N, Albaina A, Lehtiniemi M. Metabarcoding of marine zooplankton: Prospects, progress and pitfalls. J Plankton Res. 2016;38(3):393–400. https://doi.org/10.1093/plankt/fbw023
16. Bucklin A, Yeh HD, Questel JM, Richardson DE, Reese B, Copley NJ, et al. Time-series metabarcoding analysis of zooplankton diversity of the NW Atlantic continental shelf. ICES J Mar Sci. 2019;76:1162–1176. https://doi.org/10.1093/icesjms/fsz021

17. Aylagas E, Borja A, Pochon X, Zaiko A, Keeley N, Bruce K, et al. Translational molecular ecology in practice: Linking DNA-based methods to actionable marine environmental management. Sci Total Environ. 2020;774, e140780. https://doi.org/10.1016/j.scitotenv.2020.140780

18. Nunes M, Adams JB, Van Aswegen S, Macher GF. A comparison between the morphological and molecular approach to identify the benthic diatom community in the St. Lucia estuary South Africa. Afr J Mar Sci. 2019;41:429–442. https://doi.org/10.2989/1814232X.2019.1689169

19. Farrell MJ, Govender D, Hajibabaei M, Van Der Bank M, Davies TJ. Bacterial diversity in the waterholes of the Kruger National Park: An eDNA metabarcoding approach. Genome. 2019;62(3):229–242. https://doi.org/10.1139/gen-2018-0064

20. Pereira-da-Conceicoa L, Elbrecht V, Hall A, Briscoe A, Barber-James H, Price B. Metabarcoding unsorted kick-samples facilitates macroinvertebrate-based biomonitoring with increased taxonomic resolution, while outperforming environmental DNA. Environ DNA. 2020;3:353–371. https://doi.org/10.1002/edn3.116

21. Govender A, Groeneveld J, Singh S, Willows-Munro S. The design and testing of mini-barcode markers in marine lobsters. PLoS ONE. 2019;14(1), e0210492. https://doi.org/10.1371/journal.pone.0210492

22. Govender A, Singh S, Groeneveld J, Pillay S, Willows-Munro, S. Experimental validation of taxon-specific mini-barcode primers for metabarcoding of zooplankton. Ecol Appl. 2022;32(1), e02469. https://doi.org/10.1002/eap.2469

AUTHORS:
Craig D. Morris[1,2]
Kevin P. Kirkman[2]
Pete J.K. Zacharias[3]

AFFILIATIONS:
[1]Agricultural Research Council – Animal Production (ARC-AP), Pietermaritzburg, South Africa
[2]School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa
[3]Consultant in Vegetation/Animal Systems, Johannesburg, South Africa

CORRESPONDENCE TO:
Craig Morris

EMAIL:
morris@ukzn.ac.za

DATES:
Received: 26 Apr. 2022
Revised: 26 Sep. 2022
Accepted: 26 Sep. 2022
Published: 30 Nov. 2022

HOW TO CITE:
Morris CD, Kirkman KP, Zacharias PJK. Will the grass be greener on the other side of climate change? S Afr J Sci. 2022;118(11/12), Art. #13844. https://doi.org/10.17159/sajs.2022/13844

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITORS:
Jennifer Fitchett
Adriaan van der Walt

KEYWORDS:
climate change, forage quality, sourveld, mixed veld, sweetveld

FUNDING:
South African Council for Scientific and Industrial Research (CSIR)

Will the grass be greener on the other side of climate change?

Increasing atmospheric [CO_2] is stimulating photosynthesis and plant production, increasing the demand for nitrogen relative to soil supply with declining global foliar nitrogen concentrations as a consequence. The effects of such oligotrophication on the forage quality of sweetveld, mixed veld, and sourveld grasslands in South Africa, which support livestock production and native ungulates, are unknown. Soil characteristics and the herbage quality of an abundant grass are described from baseline historical (mid-1980s) data collected across a sweet-mixed-sour grassland gradient in KwaZulu-Natal. Sourveld occurred on the most acidic, dystrophic soils and exhibited a pronounced decline in leaf nitrogen, digestibility, and other macronutrients during winter, in sharp contrast to sweetveld, on nutrient-rich soils, where forage quality varied little seasonally. In a carbon-enriched, warmer, and most likely drier future climate, we predict that forage quality will not be substantially altered in sweetveld where soil nutrients and temperature are not limiting but that sourveld could become 'sourer' because soil nutrients will be inadequate to match higher plant production promoted by elevated [CO_2] and warmer and longer growing seasons. Reassessing historical data and seasonal and spatial monitoring of forage quality will enable assessment of past and future impacts of climate change on grassland forage quality.

Significance:

- Grassland forage quality will likely decline with elevated [CO_2] and warming, particularly in sourveld.
- Climate change could deepen and widen the sourveld winter forage bottleneck, necessitating greater supplementary feeding of livestock.

Introduction

Grasslands, including the C_4-dominated grasslands of South Africa, face an uncertain future in a rapidly changing climate. The ongoing rise in the atmospheric concentration of anthropogenically derived carbon dioxide (CO_2) could increase carbon sequestration, biomass production, and the water use efficiency of grasses[1,2], while also favouring woody species and alien invasive plants[3]. Reduced rainfall coupled with more frequent and severe droughts will further limit the production of herbage for livestock and the indigenous herbivores that grasslands support.[4,5] Another – largely unrecognised – threat to grazing animals posed by elevated CO_2 concentrations [CO_2] is an insidious decline in forage quality globally because of an increasing limitation of soil nitrogen (N) supply to grasses growing faster over an extended growing season in a warmer, carbon-enriched atmosphere: foliar [N] has declined by 9% globally over the last four decades.[6] Diminishing foliar [N] could cascade through ecosystems, slowing protein flow from plants to insect and mammalian herbivores.[7] Even small decreases in the protein content and digestibility of forage would adversely affect animal health, reproduction, and weight gains.[7,8]

The potential impacts of elevated [CO_2] (eCO_2) and other climate change drivers on forage quality will occur across a well-recognised and agronomically important spatiotemporal gradient in South Africa, from sweetveld through mixed veld to sourveld. Livestock on sourveld require supplementary feeds and licks for up to 6 months[9] because of the marked reduction in forage quality in autumn. This drop in quality of sourveld is caused by the translocation of foliar nutrients to roots at the end of the growing season.[10] In contrast, foliar nutrients remain sufficiently high to maintain livestock production throughout the year in sweetveld areas whereas mixed veld displays intermediary seasonal quality changes.[11] Generally, with notable exceptions[11], 'sour' grassland occurs on dystrophic soils in cool areas where high rainfall favours high primary production, but nutrient supply is limited, whereas sweetveld predominates on base-rich soils in hotter, usually lower-lying, areas, where low and erratic rainfall rather than nutrient availability restricts grass productivity[11,12]. Anthropogenic climate change could substantially alter the sweet-sour forage quality gradient because temperature, soil moisture and [CO_2] interact to determine the balance between carbon assimilation and soil nutrient availability (particularly size of the mineralisable soil N pool) that determines spatial and seasonal differences in forage productivity and quality.[12,13]

To assess whether sourveld is becoming more 'sour' and sweetveld less 'sweet' owing to eCO_2 and other climate change drivers, we present historical plant quality data (collected in 1985–1986) to (1) describe soil physicochemical and plant foliar nutrient gradients across sweet-mixed-sourveld sites, and (2) provide a baseline for detecting any oligotrophication that may have already occurred over the last third of a century. Atmospheric [CO_2] has risen by more than 20% (346 to 418 ppm) since the mid-1980s (https://gml.noaa.gov/ccgg/trends/), during which spatially variable temperature increases in mean annual temperature[14] of 0.01 °C/year to 0.03 °C/year and mean annual precipitation changes[15] ranging from minus 12 mm to positive 14 mm have been recorded over South Africa. These trends are likely to accelerate because the southern African region is a global hotspot of climate change.[16] We also consider uncertainties in the future likely trajectories of forage quality shifts in South African grasslands.

Methods

Plant quality and soil characteristics were assessed at 31 sites across a sweet-sour grassland gradient in KwaZulu-Natal (Supplementary table 1).[17] From 1985 to 1986, each site was visited at about 73-day intervals to harvest

foliar material of a consistent age – the top two leaves and a bud of vegetative tillers – from 20 plants of *Themeda triandra* (red grass), an abundant grass at all sites. We also refer to a wider study of winter (July) grass quality in the grassland biome in other provinces that used the same methods.[18] Expert knowledge and literature were used to classify sites as 'sweet', 'sour' and 'mixed' types.

Plant quality analyses[17,18] included analysis for cellulase dry matter digestibility (%) and leaf nitrogen concentration (%), as well as chemical elemental analysis (N, P, K, Ca, Mg, Na, Zn, and S) (Supplementary table 2).

Topsoil was sampled (to 200 mm depth) from 20 combined auger points and assessed for particle size (texture), organic matter, field moisture capacity, pH, exchangeable acidity, acid saturation, effective cation exchange capacity, and P, K, Ca, Mg, Na (Supplementary table 3).[17,18]

Forage quality (all seasons) and soil physicochemical gradients in KwaZulu-Natal were examined using principal component analysis of cross-correlation matrices. Seasonal differences in N% between grassland types were assessed with permutation analyses of variance (9999 permutations).

Results

There was a strong forage quality gradient along which sourveld sites were most distinct for their low leaf digestibility, N%, and cation concentrations (Figure 1). Digestibility doubled and N% ranged five-fold across this gradient (Figure 1b,c).

Sweetveld soils were less acidic, had lower organic matter and capacity to hold water, but had substantially more exchangeable cations available than soils from sourveld; mixed veld sites were intermediate (Figure 2).

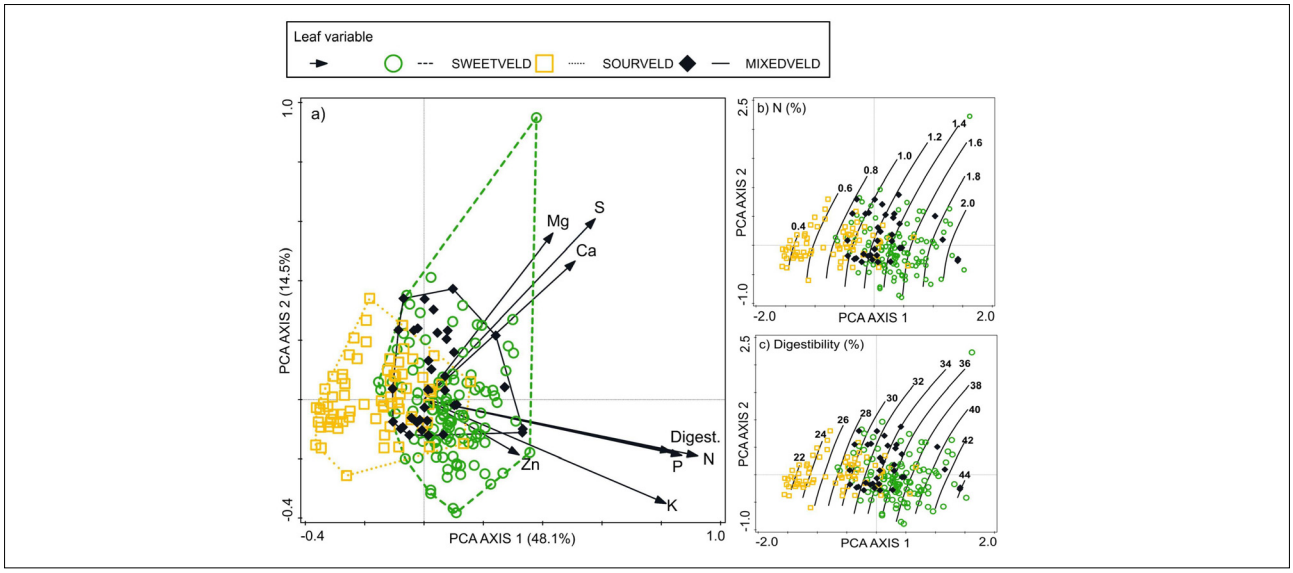
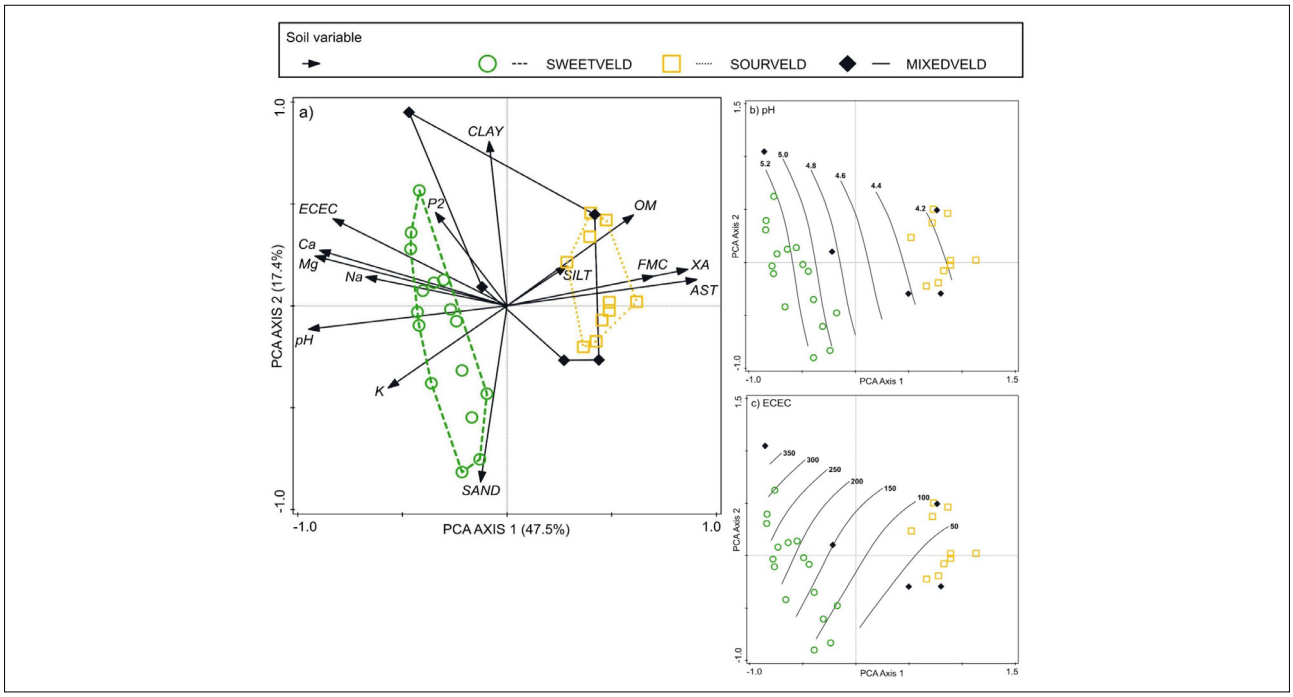


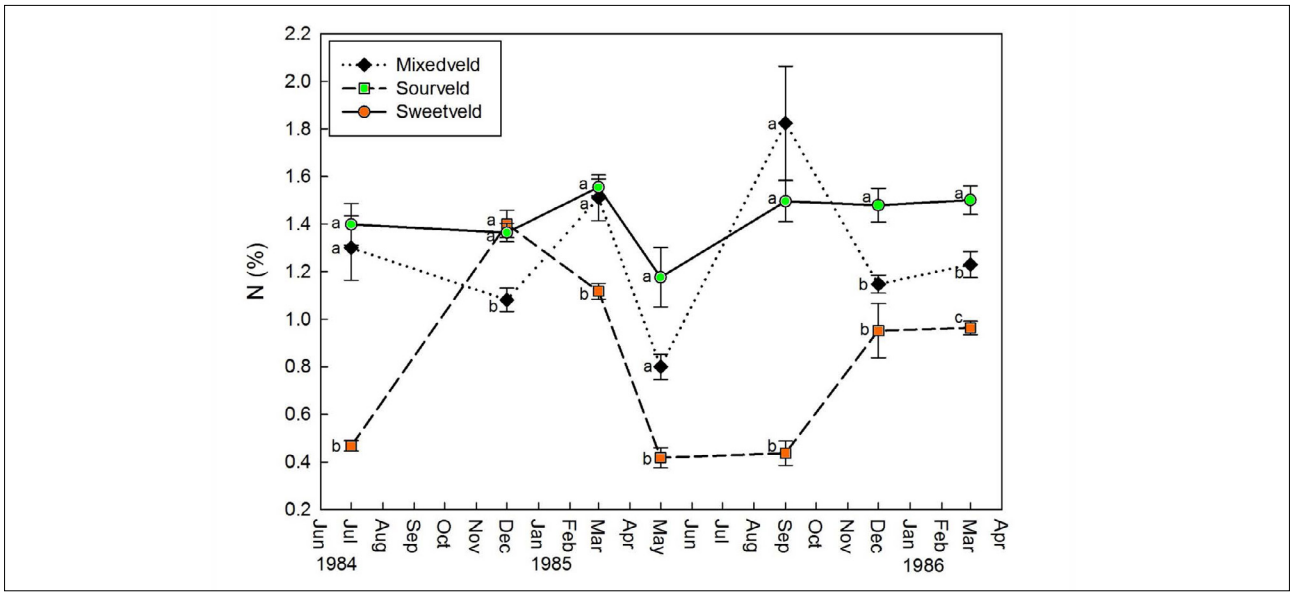
Figure 1: First two axes of a principal component analysis (PCA) of leaf variables measured at 31 locations in three grassland types in KwaZulu-Natal (a), and trend in leaf nitrogen (b) and digestibility (c) across the ordination.



AST, acid saturation; ECEC, effective cation exchange capacity, FMC, field moisture capacity; OM, organic matter; XA, exchangeable acidity

Data from Zacharias[17]

Figure 2: First two axes of a correlation-type principal component analysis (PCA) of topsoil variables measured at 31 locations in three grassland types in KwaZulu-Natal.



Data from Zacharias[17]

Figure 3: Seasonal trends in mean (±s.e.) leaf nitrogen concentration (%) measured in three grassland types in KwaZulu-Natal. Means within sampling dates with letters in common were not different ($p = 0.05$).

Sweetveld had consistently high foliar N%, declining only somewhat towards the winter of the second sampling season (Figure 3). In contrast, N concentration, similar to digestibility[17], declined to markedly low levels from late summer through to early spring in sourveld. Quality trends for mixed veld were not consistent nor pronounced. Summer and winter levels of N above 1.0% and below 0.5%, respectively for sweetveld and sourveld in KwaZulu-Natal, matched the winter extremes measured elsewhere in the grassland biome.[18] Also, in sourveld, and to a lesser extent in mixed veld, leaf concentrations of the macronutrients, P, K, Mg, and S were seasonally variable.[17]

The links between forage quality, soil characteristics and environment were weak, with increasing altitude the only consistent predictor of 'sourness' in KwaZulu-Natal.[17,18]

Discussion

Future reductions in the forage quality of sourveld and sweetveld will depend on how other major climate change drivers (temperature and precipitation) interact with the carbon fertilisation effect (CFE) to alter the balance between plant growth-driven demand for, and soil supply of, nutrients, primarily N.[6,7,12] The CFE would be most pronounced when resources and environmental conditions do not restrict plant growth.[19] However, the effects of multi-way interactions between climate drivers on plants and soils, particularly on forage quality, are poorly understood; these interactions can be complex, multiplicative[20,21], and species-specific[22]. Given these uncertainties and the current growth limitations prevailing in sour- and sweetveld[11,12], we tentatively predict the following potential shifts in plant quality (leaf [N], digestibility, and fibre content) under climate change.

Sweetveld mostly occurs in semi-arid regions, where soil moisture, not nutrients, limit plant growth.[12] Despite lower rainfall and increased evaporative demand with warming predicted for semi-arid regions[5], plant growth could increase because of more efficient water use under eCO_2 resulting from stomal closure[1,22]. Elevated temperatures combined with the CFE could stimulate grass production, lowering the N content and digestibility of herbage.[8] However, extreme and prolonged droughts and heat waves, both of which will become more frequent with climate change[5,16], will limit carbon assimilation and nutrient availability by curtailing microbial decomposition and nutrient cycling[19,20]. The consequences of climate change in semi-arid regions are still uncertain[21] but it is not likely that sweetveld will experience a consistent large directional change in productivity and forage quality in the future.

In sourveld, predominately in higher lying, cooler, moister climes, nutrient-poor soils and low temperatures act together to constrain the quality of forage during the non-growing season.[11,12] Warming, especially in spring and autumn, and eCO_2 would likely enable a longer period of growth, further limiting soil nutrient supply[6] and reducing winter forage quality. Increased N mineralisation with warming could mitigate this potential future decline in forage quality on organic mountain soils[23] but not on the dystrophic mineral soils of outlier sourveld areas at low elevations[11].

The critical winter forage bottleneck in forage quality in sourveld (Figure 3) is likely to be exacerbated in the future because eCO_2-driven reductions in protein content and digestibility in late summer and autumn occur when nutrients in senescing plants are already below critical levels for livestock.[9,24] Consequently, supplementary feeding costs will increase[8] while wild ungulates would need to forage differently to match their metabolic requirements[25]. Higher-quality C_3 grasses that remain greener for longer in autumn could obtain a competitive advantage over C_4 species in the future[26] but only at the higher and far western margins of the grassland biome, and perhaps not to any significant extent[2].

Research is required across a sweet-mixed-to-sourveld gradient to understand patterns and mechanisms of seasonal nutrient flows between plant parts – our knowledge of these is still surprisingly rudimentary.[10] Also requiring investigation are species-specific responses to interactions between multiple climate change drivers[22], the potential effects on plant growth and quality of ongoing atmospheric nitrogen deposition[20], and the extent to which CFE effects on plant quality could be modified by downregulation of photosynthesis through acclimation and progressive N limitation over time[1,7,8]. We also recommend resampling sites with historical plant data, such as those presented here (Supplementary tables 1 and 2), to establish the extent to which CFE-driven shifts in forage quality may have already occurred over the last few decades, and to regularly, widely, and seasonally monitor shifts in leaf stoichiometry (at minimum C:N ratios[7,13]) to establish the degree and extent of climate-driven oligotrophication[8].

Conclusion

Climate change has the potential to alter, in agronomically important ways, the current spatial and seasonal patterns of grass forage quality in South African grasslands. We predict the greatest 'souring' will occur in sourveld, with a minimal response in sweetveld, but there are many uncertainties as to the direction and rate of change in forage quality and the extent to which such changes will affect livestock production and

wild ungulates. Further detailed research and regular monitoring are required to assess if, where, how, and why forage quality of grasslands in South Africa is responding to climate change.

Acknowledgements

The original field studies conducted by P.J.K.Z. and K.P.K. were funded by the Co-operative Scientific Programmes (Terrestrial Ecology) of the Council for Scientific and Industrial Research (CSIR). Field, laboratory, and library assistance provided by various people during those studies is acknowledged, as are the conservation organisations and private landowners who facilitated access to sample sites.

Competing interests

We have no competing interests to declare.

Authors' contributions

C.D.M. conceptualised the study. P.J.K.Z. and K.P.K. undertook the original field studies. C.D.M. analysed the data and wrote the draft manuscript with editorial input from P.J.K.Z. and K.P.K.

References

1. Lee M, Manning P, Rist J, Power SA, Marsh C. A global comparison of grassland biomass responses to CO_2 and nitrogen enrichment. Philos Trans R Soc Lond B. 2010;365(1549):2047–2056. https://doi.org/10.1098/rstb.2010.0028
2. Wand SJ, Midgley GF, Jones MH, Curtis PS. Responses of wild C_4 and C_3 grass (Poaceae) species to elevated atmospheric CO_2 concentration: A meta-analytic test of current theories and perceptions. Glob Change Biol. 1999;5(6):723–741. https://doi.org/10.2989/10220119.2014.939996
3. O'Connor TG, Puttick JR, Hoffman MT. Bush encroachment in southern Africa: Changes and causes. Afr J Range Forage Sci. 2014;31(2):67–88. https://doi.org/10.1007/978-81-322-2265-1_7
4. Giridhar K, Samireddypalle A. Impact of climate change on forage availability for livestock. In: Sejian V, Gaughan J, Baumgard L, Prasad C, editors. Climate change impact on livestock: Adaptation and mitigation. New Delhi: Springer; 2015. p. 97–112. https://doi.org/10.1007/978-81-322-2265-1_7
5. Davis CL, Vincent K. Climate risk and vulnerability: A handbook for southern Africa. 2nd ed. Pretoria: CSIR; 2017. https://researchspace.csir.co.za/dspace/handle/10204/10066
6. Craine JM, Elmore AJ, Wang L, Aranibar J, Bauters M, Boeckx P, et al. Isotopic evidence for oligotrophication of terrestrial ecosystems. Nat Ecol Evol. 2018;2(11):1735–1744. https://doi.org/10.1038/s41559-018-0694-0
7. Mason RE, Craine JM, Lany NK, Jonard M, Ollinger SV, Groffman PM, et al. Evidence, causes, and consequences of declining nitrogen availability in terrestrial ecosystems. Science. 2022;376(6590). https://doi.org/10.1126/science.abh3767
8. Augustine DJ, Blumenthal DM, Springer TL, LeCain DR, Gunter SA, Derner JD. Elevated CO_2 induces substantial and persistent declines in forage quality irrespective of warming in mixedgrass prairie. Ecol Appl. 2018;28(3):721–735. https://doi.org/10.1002/eap.1680
9. Kirkman KP, Moore A. Perspective: Towards improved grazing management recommendations for sourveld. Afr J Range Forage Sci. 1995;12(3):135–144. https://doi.org/10.1080/10220119.1995.9647883
10. Weinmann H. Seasonal chemical changes in the roots of some South African highveld grasses. J S Afr Bot. 1940:131–145.
11. Tainton NM. The production characteristics of the main grazing lands of South Africa. In: Tainton NM, editor. Veld management in South Africa. Pietermaritzburg: University of Natal Press; 1999. p. 46–52.
12. Ellery WN, Scholes RJ, Scholes MC. The distribution of sweetveld and sourveld in South Africa's grassland biome in relation to environmental factors. Afr J Range Forage Sci. 1995;12(1):38–45. https://doi.org/10.1080/10220119.1995.9647860
13. Kunz RP, Schulze RE, Scholes RJ. An approach to modelling spatial changes of plant carbon: Nitrogen ratios in southern Africa in relation to anticipated global climate change. J Biogeogr. 1995;1(2/3):401–408. https://doi.org/10.2307/2845936
14. Jury MR. Climate trends across South Africa since 1980. Water SA. 2018;44(2):297–307. https://doi.org/10.4314/wsa.v44i2.15
15. Marumbwa FM, Cho MA, Chirwa PW. Analysis of spatio-temporal rainfall trends across southern African biomes between 1981 and 2016. Phys Chem Earth A/B/C. 2019;114(Dec), Art. #102808. https://doi.org/10.1016/j.pce.2019.10.004
16. Engelbrecht FA, Monteiro P. The IPCC Assessment Report Six Working Group 1 report and southern Africa: Reasons to take action. S Afr J Sci. 2021;117(11–12), Art. #12679. https://doi.org/10.17159/sajs.2021/12679
17. Zacharias PJ. The seasonal patterns in plant quality in various ecological zones in Natal [MSc thesis]. Pietermaritzburg: University of Natal; 1990. https://researchspace.ukzn.ac.za/handle/10413/20233
18. Kirkman KP. Factors affecting the seasonal variation of veld quality in South Africa [MSc thesis]. Pietermaritzburg: University of Natal; 1988. https://researchspace.ukzn.ac.za/handle/10413/20232
19. Seibert R, Donath TW, Moser G, Laser H, Grünhage L, Schmid T, et al. Effects of long-term CO_2 enrichment on forage quality of extensively managed temperate grassland. Agric Ecosyst Environ. 2021;312, Art. #107347. https://doi.org/10.1016/j.agee.2021.107347
20. Polley HW, Morgan JA, Fay PA. Application of a conceptual framework to interpret variability in rangeland responses to atmospheric CO_2 enrichment. J Agric Sci. 2011;149(1):1–4. https://doi.org/10.1017/S0021859610000717
21. Song J, Wan S, Piao S, Knapp AK, Classen AT, Vicca S, et al. A meta-analysis of 1,119 manipulative experiments on terrestrial carbon-cycling responses to global change. Nat Ecol Evol. 2019;3(9):1309–1320. https://doi.org/10.1038/s41559-019-0958-3
22. Pastore MA, Lee TD, Hobbie SE, Reich PB. Interactive effects of elevated CO_2, warming, reduced rainfall, and nitrogen on leaf gas exchange in five perennial grassland species. Plant Cell Environ. 2020;43(8):1862–1878. https://doi.org/10.1111/pce.13783
23. Carbutt C, Edwards TJ, Fynn RW, Beckett RP. Evidence for temperature limitation of nitrogen mineralisation in the Drakensberg Alpine Centre. S Afr J Bot. 2013;88:447–454. https://doi.org/10.1016/j.sajb.2013.09.001
24. Milchunas DG, Mosier AR, Morgan JA, LeCain DR, King JY, Nelson JA. Elevated CO_2 and defoliation effects on a shortgrass steppe: Forage quality versus quantity for ruminants. Agric Ecosyst Environ. 2005;111(1–4):166–184. https://doi.org/10.1016/j.agee.2005.06.014
25. Searle KR, Hobbs NT, Gordon IJ. It's the "foodscape", not the landscape: Using foraging behavior to make functional assessments of landscape condition. Isr J Ecol Evol. 2007;53(3–4):297–316. https://doi.org/10.1560/IJEE.53.3.297
26. Chamaillé-Jammes S, Bond WJ. Will global change improve grazing quality of grasslands? A call for a deeper understanding of the effects of shifts from C_4 to C_3 grasses for large herbivores. Oikos. 2010;119(12):1857–1861. https://doi.org/10.1111/j.1600-0706.2010.19070.x

AUTHORS:
Gregor D. Bader[1]
Manuel Will[2]

AFFILIATIONS:
[1]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany
[2]Department of Early Prehistory and Quaternary Ecology, University of Tübingen, Tübingen, Germany

CORRESPONDENCE TO:
Gregor Bader

EMAIL:
gregor.bader@uni-tuebingen.de

DATES:
Received: 07 June 2022
Revised: 08 Sep. 2022
Accepted: 14 Oct. 2022
Published: 30 Nov. 2022

HOW TO CITE:
Bader GD, Will M. 70 years later – New research at Holley Shelter, a Middle and Later Stone Age site in KwaZulu-Natal, South Africa. S Afr J Sci. 2022;118(11/12), Art. #14069. https://doi.org/10.17159/sajs.2022/14069

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☒ All data included
☐ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Jemma Finch

KEYWORDS:
Middle Stone Age, KwaZulu-Natal, palaeoenvironment, cultural evolution, human-landscape interaction

FUNDING:
DFG Project WI 4978/3-1, Heidelberg Academy of Science, Senckenberg Centre for Human Evolution and Palaeoenvironment

70 years later – New research at Holley Shelter, a Middle and Later Stone Age site in KwaZulu-Natal, South Africa

In 1952, Gordon Cramb published the first report on his excavations at Holley Shelter, in KwaZulu-Natal, South Africa. Although Cramb's work implied organic preservation associated with a unique stone tool assemblage from a Middle Stone Age (MSA) context, Holley Shelter disappeared from the archaeological landscape until 2015, when we provided a reinvestigation of the lithic assemblages from the 1950s. Our study indicated a tentative chronological framework around Marine Isotope Stage 3 (MIS3) for the human occupation of the site. However, Cramb's excavations did not produce reliable chronometric ages and the botanical and zoological finds from Holley Shelter remain for the most part unstudied. This is problematic as the site constitutes one of the few inland MSA rock shelters of KwaZulu-Natal featuring organic preservation. In 2022, 70 years after Cramb's first report, we started a new research project focusing on renewed excavations to obtain archaeological remains from a controlled stratigraphy, absolute chronometric ages and reliable data on the palaeoenvironment at the border between the coastal belt and the midlands of KwaZulu-Natal. Here, we provide initial results from the first field campaign in 2022 on the stratigraphic sequence and archaeological finds and discuss their implications for future multidisciplinary research.

Significance:

- Environmental change can have a strong impact on hunter-gatherer behaviour, migration and technological choices. It is thus crucial to contextualise archaeological material with a strong palaeoenvironmental record.
- The inland of KwaZulu-Natal, South Africa, represents an understudied region in terms of Stone Age archaeological sites and palaeoenvironmental record.
- Holley Shelter is one of the few sites in South Africa with excellent organic preservation and a deep stratigraphic record, which provide a great opportunity to investigate human-landscape interaction and technological change throughout the Middle and Later Stone Age.

Previous work

Holley Shelter on Fountainhill Estate near Wartburg, KwaZulu-Natal (KZN) (Figure 1) has been known since the 1950s due to extensive excavation activities by Gordon Cramb[1,2]. In total, Cramb excavated during five campaigns at Holley Shelter between 1951 and 1959. He published his results in two contributions to the *South African Journal of Science*, the first one in 1952 – exactly 70 years ago.

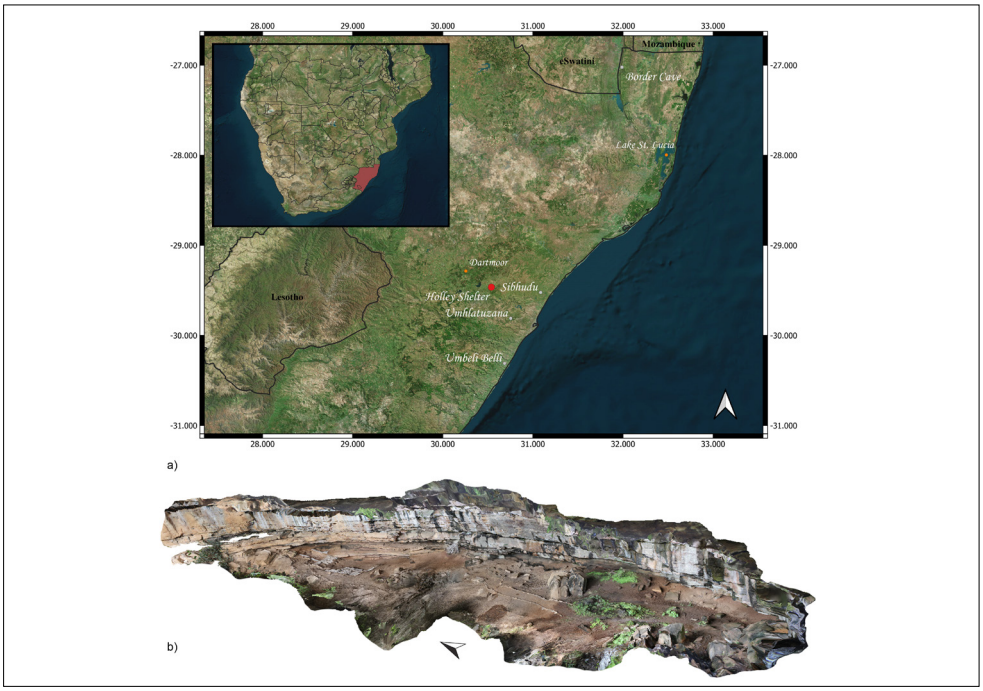


Figure 1: (a) Location of Holley Shelter and other archaeological and palaeoenvironmental sites mentioned in the text. (b) Three-dimensional model of the site produced with Agisoft Metashape Professional (by G. Bader).

In 2015, Bader et al.[3] performed a new study on the stone tool assemblages excavated by Cramb, aiming to evaluate the chrono-cultural framework of the Middle Stone Age (MSA) occupations. They recognised three distinct occupational horizons. The upper unit contained assemblages dominated by blades on hornfels with a characteristic core removal strategy, facetted platforms and frequent splintered pieces on hornfels. The middle part of the sequence was characterised by the same blade technology, fewer splintered pieces and a high number of unifacial points. The lowermost unit had lower artefact numbers dominated by quartz and bipolar percussion. When compared to lithic assemblages of the region, Holley Shelter showed similarities to the Sibudan assemblages at the MSA site of Sibhudu[4,5], indicating a potential age of 60–50 ka.

Research questions

The MSA of South Africa plays a key role in current discussions on the early cultural evolution of our species. In KZN, only six MSA rock shelter sites have been excavated, with key sites such as Sibhudu[5] and Border Cave[6]. Holley Shelter and Border Cave are the only two published inland MSA rock shelters in KZN, providing potential for good organic preservation. While Border Cave, far in the north of KZN, recently received increasing attention due to new research by Backwell and colleagues[6], Holley Shelter was last excavated in the 1950s.

The site provides the unique opportunity to gain information on the palaeoenvironment in the central hinterland of KZN well into the Pleistocene. In this area, for example, the oldest pollen record currently reaches back to about 12 000 cal BP at Dartmoor.[7] Older records about palaeolandscapes and palaeoenvironments come from Border Cave in the north and the coastal area in the east, e.g. Lake St. Lucia[8] or from Sibhudu[9]. Thus, in the course of ongoing effort to investigate past human behaviour and palaeoecology, we began new excavations at Holley Shelter in 2022 with the following research questions for our first 4-week campaign:

1. Are there intact archaeological deposits left?
2. Does the site feature a discernible stratigraphic sequence which can be dated?
3. Is there preservation of organic remains?
4. Can we confirm the existence of distinct techno/typological units in the stratigraphic sequence?

Preliminary results

Cramb excavated in two areas of the shelter. In his first campaign in 1951, he dug in the southern part, near the entrance (Figures 1 and 2), documenting a Later Stone Age (LSA) occupation at the top, and typical MSA artefacts including bone tools below.[2] Although substantial amounts of intact deposits remain, we refrained from excavating this area in our 2022 campaign due to massive rock slabs lying on the deposits. We might expand excavations to this part of the shelter in future work, as Cramb documented LSA artefacts only in this smaller area, whereas the larger area yielded exclusively MSA material. In the following campaigns, Cramb focused on this larger 'habitable area' at the northern end of the shelter.

Cramb used a square grid system in this area and painted the letters of his squares on the shelter backwall.[3] We focused on this area as it was easily accessible and had the highest chance to preserve the full sequence of archaeological deposits. Two test trenches were opened. The first at the northern end of the shelter bordering Cramb's former main excavation area, including squares 12/57, 12/58 and 12/59 (Figure 2), is also called the 'northern section'. Whereas Cramb used square yards for his excavation grid, we are using square metres and hence we can only coarsely correlate them with each other. Different from Cramb who used an alphanumeric code for his squares (e.g. X1, W2), our squares' names are purely numeric, according to their absolute coordinates (e.g. square 12/57 starts at $x = 12\,000$ and $y = 57\,000$). A second trench was sunk

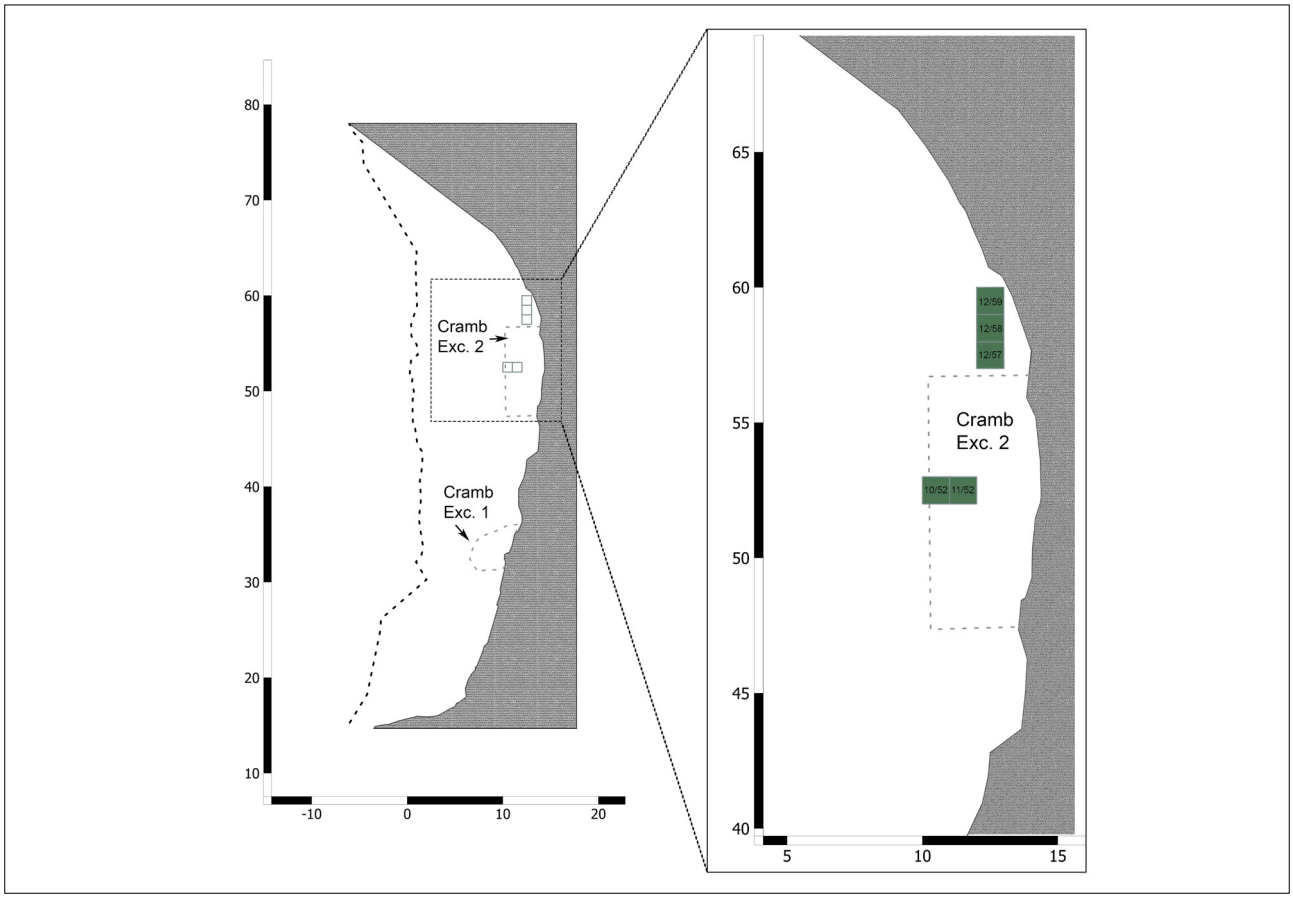


Figure 2: Excavation plan of Holley Shelter. Cramb Exc. 1 and 2 mark the areas of Cramb's excavations in the 1950s. Exc. 1 = 'smaller habitable area', Exc. 2 = 'larger habitable area'. Green squares were excavated by the TSAARU team in 2022. Figure by G. Bader.

at the western extension of Cramb's main excavation area (Figure 2). Following his field notes, he excavated only 60 cm deep without reaching bedrock there. The aim of this trench (squares 10/52 and 11/52) was to document the lower extent of these deposits below the 60-cm pit. Our layer taxonomy follows the system applied by the Sibhudu team using names from top to bottom in alphabetical order. Due to the musical preference of the excavators we decided to name the layers in honour of our favourite rock and metal bands.

Our new excavations confirm the existence of intact and stratified archeological deposits in both trenches. Table 1 provides an overview of finds measured in the field and recovered from screening of sediments. In the northern section, below a 5–10-cm thick disturbed surface horizon, we reached the first intact layer AVA (Avantasia), an orange-brown (5YR 4/4) layer of sandy silt deposits with several charcoal inclusions (Figure 3). AVA contains several small hearths, typical MSA stone artefacts and substantial amounts of faunal remains. Few of the stone tools are retouched but many of them have faceted platforms. Hornfels is the most common raw material but occasionally sandstone artefacts occur. Below AVA, layer BIB (Beast in Black) is clearly distinguishable as a grey (5YR 5/2) sandy silt (Figure 3). In the southern part, BIB contains a massive hearth expanding almost over the entire square. BIB is substantially richer in artefacts, including numerous splintered pieces similar to the ones identified in the Cramb collection (Figure 4d–f)[3], unifacial points (Figure 4a–c), frequent blades and points with faceted platforms, large amounts of faunal remains (Figure 4g–p) and big pieces of charcoal. Fossilised plant remains were also recovered. There appears to be a clear correlation between higher find densities and the hearth features, whereas the surrounding areas contain only a few artefacts.

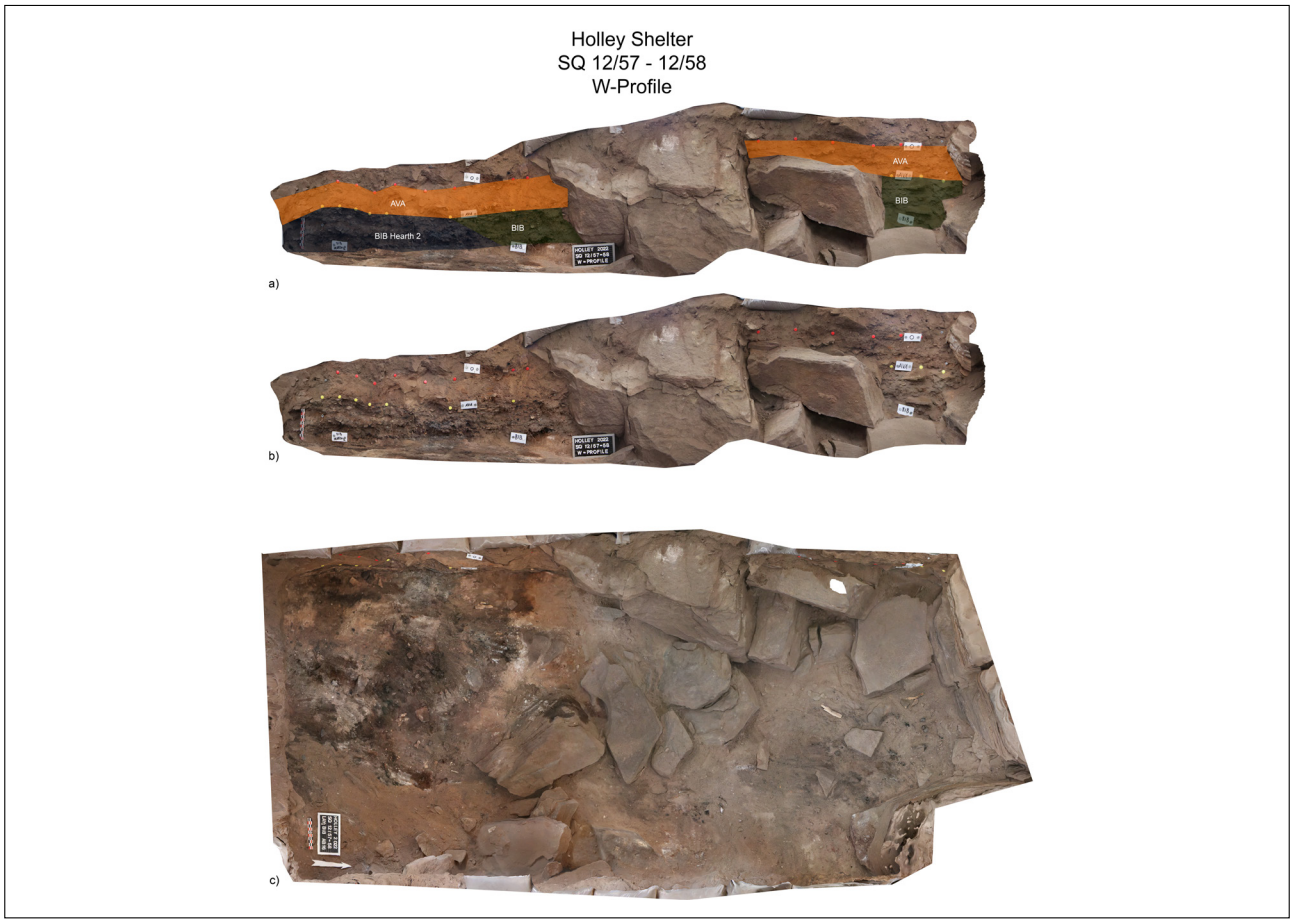


Figure 3: (a) West-profile of the northern section with layers redrawn, (b) west-profile of the northern section (original), and (c) three-dimensional model of the northern section at the end of the 2022 excavation season. Figures by G. Bader.

Table 1: Archaeological finds, per layer, from the 2022 excavation campaign at Holley Shelter

Layer	Lithics >2cm (*n*)	Lithics <2cm (*n*)	Fauna identifiable (*n*)	Fauna non-identifiable (g)	Charcoal >2cm (*n*)	Charcoal <2cm (g)	Ochre >2cm/ worked (*n*)	Ochre <2cm (g)
AVA	123	311	204	1679.9	39	85.1	4	47.5
BIB	385	521	309	2573.8	107	389.7	17	279.4
YBA	95	90	26	258.3	28	35.2	13	40.8
GYS	91	103	2	45.1	8	31.7	4	54.8
GWS	81	167	6	86.4	19	82.8	1	22.7
GBA	76	75	0	5.4	23	50.5	4	4.6
Total	**851**	**1267**	**547**	**4648.9**	**224**	**675**	**43**	**449.8**

Figure 4: Artefacts from layer BIB in the northern section at Holley Shelter: (a–c) unifacial points, (d–f) splintered pieces, (g, h, k, m, n, o) identifiable faunal remains, (i, j) bone flakes, (l, p) bones with cutmarks. Photos by G. Bader.

In the smaller trench at the bottom of the old Cramb excavation area, the first 10 cm of the deposits represent backfilling and disturbance over the past 60–70 years. The sediments below, however, are intact. Because we cannot yet correlate the layer system from the northern section with the Cramb area, we decided to use descriptive layer names in this trench which will be replaced by alphabetical names once we can connect both trenches. The first intact layer in the Cramb area is yellow brown ash (YBA), a hearth feature containing substantial amounts of hornfels artefacts, faunal and botanical remains and frequent charcoal. The top surface of YBA lies approximately 1.20 m below the top surface of AVA in the northern section. Underneath YBA follow layers grey yellow speckled (GYS), grey white speckled (GWS) and grey brown ash (GBA). GWS also features well-preserved, small hearths. Stone artefacts in those lower layers are typically knapped from sandstone and quartz, the latter often flaked by bipolar technique. Although bone preservation is poor, charcoal is still preserved and there is good evidence for further botanical remains.

A further important observation concerns the distribution of stone artefact categories. In 2015 we suggested that Cramb's original assemblage was likely biased due to high numbers of modified pieces and low numbers of small debitage in the collection. However, our new excavations so far confirm this trend. A total of 851 single finds >2 cm are opposed to 1267 pieces of small debitage <2 cm of which only 182 are <1 cm. At the same time, the upper horizons AVA and BIB as well as YBA contain considerable amounts of tools ranging between 6.3% and 9.4% of the lithic assemblage (including splintered pieces which are the

focus of more detailed residue and microwear studies to ascertain their function). In AVA and BIB, splintered pieces and unifacial points are most frequent, as in the old collections. In contrast, the lowermost horizons show only 0–1% modified tools and lack retouched points or splintered pieces. In general, the find densities of lithics are comparatively low but variable, with the lowest value in AVA (n=1555/m^3) and the highest values in BIB (n=3682.9/m^3) and GWS (n=4829.4/m^3).

Discussion

The new excavations at Holley Shelter revealed a well-stratified archaeological sequence with excellent organic preservation and clearly delineated anthropogenic hearth features. Our preliminary results confirm earlier observations of occupational horizons with different techno-typological characteristics. AVA and BIB correlate well with Cramb's Inches 0–6 and 6–12, exhibiting large numbers of hornfels flakes and blades with faceted platforms and splintered pieces. GYS, GWS and GBA overlap with Cramb's lower Inches 30–36 and 36–42, having little to no hornfels artefacts, lower find numbers in general, bipolar percussion on quartz nodules and rare retouched tools.

The low number and proportion of waste products (small debitage) at Holley Shelter stand in sharp contrast to MSA sites closer to the coast in KZN such as Sibhudu or Umbeli Belli, which suggests comparatively little on-site reduction of raw materials and – in combination with higher find densities being associated directly with clearly delineated hearth features – potentially indicates more small-scale and short-term stays. This would imply a different site function for Holley Shelter compared to the more residential MSA sites closer to the coast – a hypothesis which requires further testing.

Future perspectives

Many research questions at Holley Shelter remain active targets for future work in the coming years. In order to provide an absolute chronological framework, we took samples for both optically stimulated luminescence and ^{14}C dating from the sequence. Due to excellent organic preservation, we intend to establish a comprehensive archive for palaeoenvironmental and climate reconstructions reaching back in time much further than previously possible in the central part of KZN. We therefore created a research network with South African and European researchers focusing on pollen, phytolith, macro-botanical, faunal, charcoal, and geoarchaeological analyses – with additional studies on lithics and ochre – all currently underway. The new excavations at Holley Shelter have the potential to provide a novel perspective and essential contribution towards a broader understanding of hunter-gatherer behaviour and palaeoenvironment during the MSA, providing a bridge in time and space to the well-known research areas along the coast in the east (Sibhudu, Umbeli Belli, Umhlatuzana), the highlands of Lesotho in the west (Sehonghong, Melikane) and the mountainous north (Border Cave, Sibebe). Complementary work by our team on several other rock shelter and open-air sites in KZN and Eswatini intends to fill gaps in our knowledge and provide a regional perspective on human material culture, exchange networks and palaeoecology throughout the Pleistocene in southeastern South Africa.

Acknowledgements

We express our gratitude for the enthusiastic support of Ed Gevers and the members of the Taeuber trust at Fountainhill Estate in the uMshwati district, KZN. We are especially grateful to AMAFA for providing us with the research permit (3450) to excavate at Holley Shelter. We thank our excavation crew members Matthias Blessing, Leah Böttger, Marina-Zafra Granados, Gunther Möller and Nina Stahl for their hard work and dedication. Finally, we thank Peter Cramb who approached us so many years ago and brought our attention to Holley Shelter as the first step into this new adventure. This project was funded by the DFG Project WI 4978/3-1, the ROCEEH Project of the Heidelberg Academy of Science and the Senckenberg Centre for Human Evolution and Palaeoenvironment.

Competing interests

We have no competing interests to declare.

Authors' contributions

Both authors designed the study, collected and processed the data, and wrote the manuscript.

References

1. Cramb G. Second report on work at the Holley Shelter. S Afr J Sci. 1961;57(2):45–48.
2. Cramb G. Middle Stone Age industry from a Natal rock shelter. S Afr J Sci. 1952;48(6):181–186.
3. Bader GD, Will M, Conard NJ. The lithic technology of Holley Shelter, KwaZulu-Natal and its place within the MSA of southern Africa. S Afr Archaeol Bull. 2015;70(202):149–165.
4. Conard NJ, Porraz G, Wadley L. What is in a name? Characterizing the Post-Howiesons Poort at Sibudu. S Afr Archaeol Bull. 2012;67(196):180–199.
5. Will M, Bader GD, Conard NJ. Characterizing the Late Pleistocene MSA lithic technology of Sibudu, KwaZulu-Natal, South Africa. PLoS ONE. 2014;9(5), e98359. https://doi.org/10.1371/journal.pone.0098359
6. Backwell LR, d'Errico F, Banks WE, De la Peña P, Sievers C, Stratford D, et al. New excavations at Border Cave, KwaZulu-Natal, South Africa. J Field Archaeol. 2018;43(6):417–436. https://doi.org/10.1080/00934690.2018.1504544
7. Geer A. Holocene environmental changes derived from a montane pollen sequence in the KwaZulu-Natal midlands, South Africa [master's dissertation]. Durban: University of KwaZulu-Natal; 2013. http://hdl.handle.net/10413/11473
8. Finch JM, Hill TR. A late Quaternary pollen sequence from Mfabeni Peatland, South Africa: Reconstructing forest history in Maputaland. Quat Res. 2008;70(3):442–450. https://doi.org/10.1016/j.yqres.2008.07.003
9. Sievers C. Seeds from the Middle Stone Age layers at Sibudu Cave. South African Humanit. 2006;18(1):203–222.