South African Journal of Science
120 years
volume 120
number 11/12

Donating using blockchain-based vouchers

Heat health perceptions in minibus taxis

Oldest ophiuroid fossils in South Africa

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Volume 120
Number 11/12
November/December 2024

South African Journal of Science
120 years

eISSN: 1996-7489

Leader

Impact
Leslie Swartz 1

Profile

Ahead of his TIME? Professor Tulio de Oliveira's second nomination as one of the world's 100 most influential individuals, and he is far from done
Mathys J. Redelinghuys 2

Obituary

Brian O'Connell (1947–2024): A visionary, inclusive and caring educationist
Ramashwar Bharuthram 4

Book Reviews

Notes from the body: An autobiographical review
Helen Moffett 6

Table Mountain: Reading the rocks on an unimaginable scale
Saul Dubow 8

The country ahead of us, the country behind: A review of 'Contested Karoo'
Emma Archer 10

Modernity and its discontents: Tradition and the problem of liberation in postcolonial Africa
Veeran Naicker 12

Beyond 'publish or perish': Unpacking the complexities of academic publishing in Africa
Lily Kpobi 14

News

A tale of three Davids
David J.N. Limebeer 16

Commentaries

Transitioning to climate-resilient development pathways in South Africa: What does it take?
Anna Taylor 18

South Africa's oldest ophiuroid fossils provide rare insights into the origination of the Malvinoxhosan Realm
Robert W. Gess, Caitlin Reddy 21

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
diacriTech Technologies

Correspondence and enquiries
sajs@assaf.org.za

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

Response to Moultrie and Dorrington (2024): 'Problems and concerns with the 2022 South African census'
Nthabiseng Mkhatha, George Djolov, Diego Iturralde 25

Rejoinder to Mkhatha et al. (2024): Problems and concerns with the 2022 South African census
Tom A. Moultrie, Rob Dorrington 28

Structured Conversation

Speaking evidence to power? An interdisciplinary conversation
Mehita Iqani, Anna Feigenbaum, Laila Asmal, Linsey McGoey, A.M. Kanngieser, Alexander Dunlap, Thema Monroe-White 30

Review Article

A thematic analysis of South African opinions about COVID-19 vaccination on Twitter
Philippa Kerr, Kevin Durrheim, Maria Schuld, Davide Morselli 36

Research Articles

Assessing heat-related health perceptions in the minibus taxi industry in Tshwane, South Africa
Caradee Y. Wright, Thandi Kapwata, Nomfundo Mahlangeni, Natasha Naidoo, Candice Webster 42

Mobile phone based laser scanning as a low-cost alternative for multidisciplinary data collection
Kian J. Jeftha, Moreblessings Shoko 51

Voucher tokenisation using blockchain and smart contracts to support people in need
Tricia Harraway, James Bekker 58

Significance of international life cycle data in South African extended producer responsibility
Edwin E. Botha, Kevin G. Harding 66

Ambient seismic noise imaging of a tailings dam internal structure
Thulisile Kunjwa, Moctar Doucouré 74

Investigation of nanomaterial and hazardous emissions at coal-fired power stations in Mpumalanga, South Africa
Simphiwe Nundze, Adeniyi Ogunlaja, Kirstie Eastwood, Melusi Thwala, Paula Melariri 83

Research Letter

Late Pleistocene and Holocene fauna from Waterfall Bluff Rock Shelter, Mpondoland, South Africa
Sandee Oster, Jerome P. Reynard, Hayley C. Cawthra, Irene Esteban, Justin Pargeter, Erich C. Fisher 92

Discussions on Service Delivery

Leader

Thinking about 'service delivery'
Leslie Swartz 97

Perspective

Unwilling or unable? A critical reflection on the state of municipal water services, 2019–2024
Lubabalo Luyaba, Pilate Moyo, Nonjabulo Mbhele, Mukundi Mochotlhoane 98

Commentaries

Who will free us from 'service delivery'? Why a supposed solution is a problem
Steven Friedman 105

When law struggles to deliver: Reflections on service delivery law reform in South Africa, 1996–2024
Anél du Plessis, Oliver N. Fuo 108

Data analysis is part of a better feedback loop for local government
Gordon Inggs, Delyno du Toit 112

Reimagining local government service delivery through the quadruple helix
Maréve Biljohn, Grey Magaiza 117

Moving beyond basic service delivery for inclusive reliable infrastructure
Gillian Maree, Samkeliswe Khanyile 120

Water insecurity in Johannesburg (and beyond?)
Craig Sheridan 125

What a mess: Rethinking municipal waste management 30 years into South African democracy
Marc Kalina, Catherina Schenck 128

Navigating the complexities of long-term care service delivery for older people in South Africa
Naomi Hlongwane 132

Stuck in old ways: Towards transdisciplinary training for healthcare professionals in early intervention
Alecia Samuels 136

District health service delivery and the contribution of family physicians
Robert Mash, Jenny Nash 141

From the burden of disease to the disease of burden
Jacqueline Hoare, Heidi Matisonn 145

Cover caption
The development of a secure blockchain-based architecture for cashless donations through a digital voucher management system is described by Harroway and Bekker.

Impact

As we reach the end of this celebration year for the *South African Journal of Science*, we have received many congratulations from readers and contributors about our Journal. We thank everyone who celebrates with us.

The *SAJS* came into being before the Union of South Africa. Those of us who work for the Journal cannot ourselves claim any special credit for the unbroken 120-year history of the Journal, but we do have a responsibility to take into account what it means to be part of this history. We are also very aware of our privilege as the flagship journal of the Academy of Science of South Africa (ASSAf), funded through the Department of Science, Technology and Innovation – most South African journals do not have this luxury, or the kind of infrastructure which enables us to do what we do.

When we examine our recently published special issue celebrating our 120 years as a journal, and especially the central contribution by Jane Carruthers, historian of science and immediate past Editor-in-Chief, we can see strong evidence of many tensions in our history. The founding of any academic journal is not just an academic milestone in the narrow sense, but also a statement of identity and a projection about the future. Regardless of discipline or range of disciplines, founding a journal instantiates an imagined community and is part of the construction of that community. In our Journal's case, the identity work, conscious or otherwise, has been and continues to be centred around the central questions of what science is in South Africa and who the scientists are. Early on, it was important to establish from South Africa that there was indeed a community of scientists doing good research, and, although the audience of the Journal was partly, or even largely, South African, there was inevitably a gaze northwards to what was perceived to be the centre of scientific excellence. At the same time, and quite early in the history of the Journal, as Carruthers shows, many believed that the primary function of the Journal was to create, and maintain, a sense of home and identity for South African science and scientists. There was also, in the early history of the Journal, less uniformity than we have today in the style of presentation of scientific ideas, and a less clear distinction between 'amateur' and 'professional' scientists.

It is something to celebrate that our journal, along with others, became more professionalised, and that what readers see today in each issue is a set of contributions which have gone through an exacting process of assessment and copy-editing to produce a product with certain standard features. Our Journal does well when assessed against local and internationally used metrics, although we are aware of the limitations of these. Standard measures of journal strength and impact tell a story, but they are not designed to measure other features of a national journal which we value – features which, in the main, are more difficult to measure. Questions with which we continue to grapple include:

- To what extent is our science inclusive, and are we working harder to ensure that all aspects of the scientific process have a wide range of voices and expertise?
- What are the best ways to assess critically the extent to which our work has a positive and lasting impact?
- Whom do we leave out when we do scientific work and assess its impact?

These questions are by no means new for South African scientists, and many teams are addressing them in a range of ways, as a perusal of our Journal would show. But just as there are limitations of using metrics like Journal Impact Factor as a proxy for the quality of journals, especially interdisciplinary journals in the Global South, there are challenges in assessing broader societal impact. Recently we published a special issue (Vol. 120 No. 9/10) organised around questions of science engagement, what this means and how it can be assessed. We are very grateful to the guest editors and contributors to this volume, and to many other authors who have tackled complex and 'wicked' problems in research reported in our Journal.

Despite these useful contributions, at this time of reflection on the long past and the future of our Journal, we are concerned that we may not have done enough as yet to stimulate and support robust discussions – not just of examples of work which has been shown to have impact, important though this is, but also discussions on the ways in which, from a wide range of perspectives, we should think about and assess impact critically and constructively. As soon as we begin to ask these questions at quite a general level, we are confronted with the realities of both how important scientists are to thinking about these issues, and how constrained scientists often are in what can be achieved in the world on an ongoing basis. We need to understand not only the technicalities of our work but also the social and political context for our work and for how we engage with a range of actors, most of whom have little or no scientific training but who may have skills and expertise to make things happen, skills which we ourselves may lack.

As part of ASSAf and as part of the broader scientific community, we are in the age of intensive work and thinking about open science. At our Journal, we want to do as much as we can to support debate on open science and on questions of inclusion and science for all. This is a key part of our current identity as a journal, and we welcome strong and provocative contributions on this subject – we need, together, to strengthen our thinking on broadening the impact of science, building on the excellence that already exists.

HOW TO CITE:
Swartz L. Impact. S Afr J Sci. 2024;120(11/12), Art. #20398. https://doi.org/10.17159/sajs.2024/20398

AUTHOR:
Mathys J. Redelinghuys[1,2]

AFFILIATIONS:
[1]Centre for HIV and STIs, National Institute for Communicable Diseases, National Health Laboratory Services, South Africa
[2]Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Mathys Redelinghuys

EMAIL:
ShaneR@nicd.ac.za

HOW TO CITE:
Redelinghuys MJ. Ahead of his TIME? Professor Tulio de Oliveira's second nomination as one of the world's 100 most influential individuals, and he is far from done. S Afr J Sci. 2024;120(11/12), Art. #20390. https://doi.org/10.17159/sajs.2024/20390

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
TIME100 Health list, Tulio de Oliveira, bioinformaticist, teamwork, viral outbreaks, CLIMADE

PUBLISHED:
04 December 2024

Ahead of his TIME? Professor Tulio de Oliveira's second nomination as one of the world's 100 most influential individuals, and he is far from done

An *h*-index of 81, affiliations with several top global research institutes, more than 300 publications in esteemed journals and over 20 years of studying viral outbreaks. One might think that accolades like these would automatically bump you up the ranks of *TIME Magazine*'s list of the 100 most influential people globally. Sure, they will no doubt help along the way, but, according to Professor Tulio de Oliveira, it's in the *small* things. Think consistency, resilience, a robust core research team, and, viruses. Prof. de Oliveira, a leading genomics and bioinformatics scientist, has been nominated for the second time by *TIME Magazine*; first in the 2022 TIME100 list of the world's most influential people, and again in 2024 in its inaugural TIME100 Health list. I had the privilege of meeting one-on-one with Prof. de Oliveira (albeit online) to talk about his nominations and what they mean in the grander scheme of things.

It all started when, as a young boy, his curiosity in the wonders of nature was sparked by his love for outdoor activities like snorkelling. At the age of 10 he started coding, which would turn out to be the inception of his career as a bioinformaticist. He started studying molecular biology in Brazil and completed his studies in South Africa with a PhD in bioinformatics. To economise on time and efforts, he models any novel research he would pursue on existing models that have been well established for research in other infectious diseases and outbreak investigations. This approach stood him in good stead when the COVID-19 pandemic hit, and his outstanding research led to his first *TIME Magazine* nomination. In 2020, Prof. de Oliveira headed the group that discovered the Beta variant of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the virus responsible for causing COVID-19. Also, in 2021, he and his team confirmed the emergence of yet another SARS-CoV-2 variant that was later called Omicron. Although the world decided to impose travel restrictions on South Africa with this discovery, the swift detection of these variants was a testament to the extensive infrastructure and unmatched investment of human capital and financial resources that have gone into infectious diseases research in South Africa, in particular for HIV and tuberculosis (TB) research, that could be repurposed for the investigation of emerging diseases and outbreaks.

Many laypeople were astounded at, and somewhat suspicious of, the speed at which thousands of samples were collected, sequenced and analysed to track the evolution of the rampant SARS-CoV-2 virus. But for Prof. de Oliveira and his team, it was just another day in the laboratory: leveraging the systems and networks in place for HIV and TB, and pivoting for the COVID-19 response. He described how focal the intersection between the private hospital network and diagnostic arena in South Africa was during the COVID-19 pandemic. A strong team at ground level, comprising scientists, doctors, nurses and other supporting staff, ensured a coordinated response to a new threat. This very organised network is what helped South Africa to promptly identify and track the ever-changing virus in a seemingly chaotic healthcare landscape. South Africa's response to the COVID-19 pandemic might not have been perfect, but the pooling of public and private health resources, combined with the knowledge and skills of leading international researchers like Prof. de Oliveira, ensured a relatively smooth transition to the other side of the pandemic.

I was curious as to why a world-renowned scientist like Prof. de Oliveira chose to do his research in South Africa. His answer was quick and is simple: "Why not? We conduct the highest level of science in Africa, and our researchers are unapologetic and don't find excuses for what we can't do (yet). We make do." The firm serenity with which he spoke these words was reminiscent of the David Diop poem "…that is your Africa springing up anew… springing up patiently, obstinately…." His career recognition is affording him the opportunity to do a lot more than before, to train more African scientists and to conduct more research that is relevant to Africa. It helps improve his laboratory's competitiveness and they can therefore apply for more and larger external grants. For example, their funding has in turn allowed them to fund several large fellowships to help with capacity building, with more than one researcher supported to attend teaching and training workshops. Better funding also ultimately means a more direct response at the site where the outbreak is occurring: ensuring the availability of reagents and redirecting resources to where they are needed most. When the COVID-19 pandemic started fading, he launched an ambitious new project called the Climate Amplified Diseases and Epidemics (CLIMADE) consortium in collaboration with the Wellcome Sanger Institute. CLIMADE is the first global consortium that is led by researchers from the Southern Hemisphere to help fight a multitude of diseases that are aggravated by climate change, such as the Zika virus and cholera. The three leads of CLIMADE are from Australia, Brazil and South Africa, with the high-income nation of Australia providing pivotal support in CLIMADE's mission and vision. The approach of CLIMADE is holistic and entails capacity building in different countries, with additional support for testing and reacting locally to any imminent threat.

Prof. de Oliveira happily acknowledges that his illustrious career is supported by strong teams, and learning to work in teams is something he cannot emphasise enough. He advises young researchers like MSc and PhD students to participate in a variety of research projects wherever possible. To ensure he has a strong team that will increase the likelihood of success of any project and minimise problems, he looks for good mentors, judges a scientist by output and not personality, gives more attention to high performers, does not incentivise bad behaviour and has a core

team of 5–10 people. Some of the specific lessons he has learned in his career include: (1) going down the hierarchy and respecting the people who do the grunt work, such as people in the diagnostic laboratory – these are the people who are in the thick of it and you might learn more from them than elsewhere; (2) understanding the science, and as the director, having to do things you do not like or do the things others do not want to do; (3) hiring someone better than you, as this might free up your time for things you actually like to do – it is a team effort after all; (4) learning from your critics, asking yourself whether their comments and critique are true or not – to discern between work and personality; and (5) accepting that research is a game of persistence – he once went through 23 consecutive grant application rejections, but this gave him the opportunity to recycle and build upon them for better ideas and money.

"When the work pressure doubles, double up on everything else in your life, including your mental health." As a Zen Buddhist, he applies this logic to the amount of sleep he gets, family time, hiking and meditation. He underscores that this is not a luxury, but a necessity.

I conclude by thanking Prof. de Oliveira for availing himself for the interview (and for the personal mentoring session), and leave you with his words: "Remember that health is about people's livelihoods."

Professor Tulio de Oliveira

AUTHOR:
Ramashwar Bharuthram[1]

AFFILIATION:
[1]Emeritus Professor, Department of Physics and Astronomy, University of the Western Cape, Cape Town, South Africa

CORRESPONDENCE TO:
Ramashwar Bharuthram

EMAIL:
rbharuthram@uwc.ac.za

HOW TO CITE:
Bharuthram R. Brian O'Connell (1947–2024): A visionary, inclusive and caring educationist. S Afr J Sci. 2024;120(11/12), Art. #20285. https://doi.org/10.17159/sajs.2024/20285

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
04 December 2024

Brian O'Connell (1947–2024): A visionary, inclusive and caring educationist

Brian O'Connell started his career in 1970 as a teacher at Florida Secondary School, Ravensmead in the Cape Flats, 10 years later rising to the position of Principal of Kleinvlei Secondary School. With a deep-seated interest in tertiary education, he moved to the University of the Western Cape (UWC) as Senior Lecturer in the Faculty of Education where he served from July 1985 to March 1988, then moving to the position of Rector of Athlone College of Education. In April 1991, he joined Pensinsula Technikon as the Director of the School of Education.

In October 1995, Brian O' Connell was appointed head of the Western Cape Education Department (WCED) as Superintendent General of Education in the Province. His primary goal was to elevate the provincial education system to fully serve the people of the Western Cape, especially those who were severely disadvantaged and impoverished during apartheid. In carrying out his mandate, he experienced many challenges. None was as large as the task of communicating to the educators in the schools the possibility of job losses. The rationalisation into a unitary education system within democratic South Africa's transformation and redress agenda pointed to job losses for some teachers. As the Western Cape was better resourced than most of the other provinces in the country, the impact was expected to be more severe therein than in others. How does one communicate such distressful news to the teachers and the wider community? Brian O'Connell was a strong proponent of inclusivity and engagement. He held 24 meetings across the Province with teachers, school governing bodies (SGBs) and the wider community. He wanted to engage with them to unpack the reality of the situation, despite its harshness, motivated by the need for sense-making and understanding. Not unexpectedly, the atmosphere at the meetings was hostile and he was verbally attacked in his capacity as the messenger of unpleasant news. But the truth of what was to soon happen had to be shared with the community; his conscience would not allow him to do otherwise. Through these meetings he was able to get the communities to understand the actions that were to follow and the need for them.

In his capacity as Superintendent General of Education, his strategy was to empower the educators within the Province. He focused on developing the leadership skills of school principals and their management teams, as well as the SGBs so that they would take ownership of their schools and lead them onto successful growth paths. He introduced a shift from centralised responsibility to shared responsibility within a devolved management model. This approach, while welcomed by the school principals and SGBs, unfortunately led to a fierce conflict with the MEC for Education in the Western Cape. Not prepared to compromise his stand on inclusivity and community development, Brian O'Connell decided to give up his post, explore other opportunities or go into retirement.

At that point in time, he was encouraged by a staff member of UWC to apply for the then vacant post of Rector and Vice Chancellor. He did so, was interviewed and offered the post. In November 2001, Brain O'Connell entered as Rector and Vice Chancellor of an institution in dire straits, confronted by a crippling financial crisis including a huge debt, a disillusioned and demoralised staff complement still reeling from the trauma of retrenchments, coupled with an academic project facing collapse as student numbers dwindled by a third to less than 10 000. O'Connell faced an institution with an uncertain future. With a unique sense of care and compassion, self-belief and unflinching hope for the future, he presented a bold but carefully constructed vision for the revitalisation of UWC and emphasised the importance of its role in South Africa's reconstruction and development. Such was the vision and his humble yet strong leadership style, that he was able to galvanise all UWC stakeholders into believing that a better future was indeed possible through a collective effort. With the support of his Executive and middle managers and students and staff across the Institution, O'Connell initiated an open and transparent process towards strategically rebuilding UWC.

A short while after Professor O'Connell took office, a national working group, established by the Minister of Education to advise on the restructuring of higher education in South Africa as part of the necessary rationalisation within our new democratic order, recommended that UWC be merged with Peninsula Technikon. Not unexpectedly, the response from staff was to embark on protest action. O'Connell convinced them that an intellectual response to the possible merger was the route to follow. He responded to the ministry by presenting his vision for UWC to state officials, accompanied by a strong argument for why it was necessary for UWC to remain as a stand-alone institution for the country's transformation and redress programme as well as for its socio-economic development agenda. Government was persuaded and UWC did not merge with Peninsula Technikon. The focus now was to realise his vision in transforming UWC into a financially viable academic institution underpinned by excellence and impact in teaching and learning, research and innovation, and community engagement, that would revitalise the University's promise to work towards a post-apartheid future.

Professor O'Connell had a deep understanding of the daunting challenges facing South Africa's fledgling democracy and the importance of the role of universities as the originators of new knowledge and thought processes. He constantly articulated the need to make sense of the situation as depicted in the picture below, which he used consistently to remind the UWC family of the institution's greater purpose.

The challenge was to find the middle road that would address the dire needs of the 85% of our country's population who had previously been severely disadvantaged, while at the same time not putting fear into the 15% who were advantaged during apartheid.

When Professor O'Connell retired at the end of 2014, just 13 years after taking office, UWC had exceeded expectations. An expansive scholarly project placed it at the cutting edge of debates about social transformation. Its financial

Source: Bharuthram et al.[1] (reproduced with permission)

The O'Connell Sense-Making Scenario: The divide between the advantaged (15%) and disadvantaged (85%).

well-being was restored and the rewards of years of committed attention to the task of reorienting the University towards the task of consolidating post-apartheid freedom helped to rebuild the teaching and research capacity of the University. Within a relatively short period, it was recognised as one of the leading research-led universities in South Africa and the continent. This was confirmed by the 2015 *Times Higher Education* ranking of African universities that placed UWC 5th in South Africa and 7th in Africa. The academic landscape of UWC had dramatically transformed with an expanding set of scholarly undergraduate and postgraduate programmes accompanied by the building of modern infrastructure, which served a student population that grew from just under 10 000 in 2001 to 22 000 in 2014. UWC was described as a "shining star" by representatives of the state and members of public institutions when reflecting on what is possible with visionary leadership.

Professor O'Connell's commitment to "empowering our people" went beyond academia. He also took a bold stance in combating ignorance regarding HIV/AIDS, regularly engaging communities and their organisation on the importance of disciplined behaviour. The immense passion with which he did this earned him the role of chairperson of Higher Education South Africa's Strategic Advisory Committee on HIV/AIDS.

While his achievements at UWC would be judged by many as the highlight of his 44-year-long career, his deep commitment to community development and his overall contribution to education in South Africa as a teacher, lecturer, school principal, Rector of a College of Education, Superintendent General of the Provincial Education Department, to Rector and Vice-Chancellor of the University of the Western Cape over a period from 1970 to 2014, did not go unnoticed.

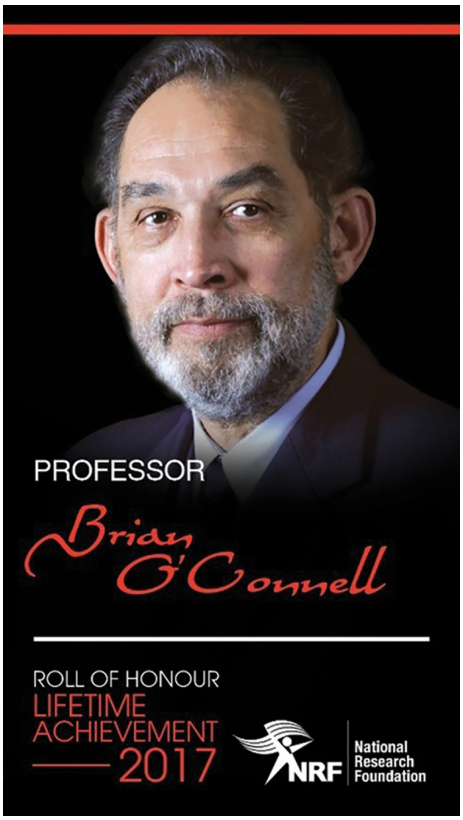


Image: Bharuthram et al.[1] (reproduced with permission)

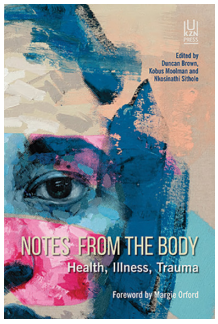
Professor Brian O'Connell received a Lifetime Achievement Award from the National Research Foundation in 2017.

Apart from receiving several honorary doctorates, Professor O'Connell was bestowed an order of knighthood (Commander of the Order of Leopold II) by her Royal Highness Princess Astrid of Belgium in 2013 for his contribution to higher education in the global space. In recognition of his immense contribution to education and development in South Africa, "driven by a passionate commitment to development in South Africa and a deep desire to make a contribution to that development", in 2017, Brian O'Connell received a Lifetime Achievement Award from the National Research Foundation.

Reference

1. Bharuthram R, Pokpas L, editors. From hope to action through knowledge: The renaissance of the University of the Western Cape, 2001–2014. Cape Town: UWC Press; 2020.

BOOK TITLE:
Notes from the body: Health, illness, trauma



AUTHORS:
Duncan Brown, Kobus Moolman and Nkosinathi Sithole

ISBN:
9781869145323 (softcover, 272pp)

PUBLISHER:
University of KwaZulu-Natal Press, Pietermaritzburg; ZAR250

PUBLISHED:
2023

REVIEWER:
Helen Moffett[1]

AFFILIATION:
[1]Independent scholar currently affiliated with Stellenbosch Institute for Advanced Study (STIAS), Stellenbosch, South Africa

EMAIL:
hecate@iafrica.com

HOW TO CITE:
Moffett H. Notes from the body: An autobiographical review. S Afr J Sci. 2024;120(11/12), Art. #18895. https://doi.org/10.17159/sajs.2024/18895

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
04 December 2024

Notes from the body: An autobiographical review

An eye gazes out from the front cover, a fragment of a portrait both sombre and colourful, by Lionel Smit. That gaze both confronts and evades the viewer, looking inward as well as outward. It is an excellent visual cue to what lies between the covers: a series of essays, poems, slices and splices of memoir, meditations and dialogues written by 21 scholars, academics, poets, composers, teachers and other creatives.

Hosted at the University of the Western Cape and funded by the Mellon Foundation, this project began, with the kind of enormous irony that would be implausible in fiction, with the intention of reflecting on the theme 'this mortal body' – a topic promptly exploded by the arrival scant months later of an invisible virus that was to reshape the world as we knew it. The planned seminars went ahead virtually, a form of disembodiment forced by the pandemic. This necessity for virtual conversations and collaboration confronted us with hard truths about how we viewed bodies, especially 'different' bodies: those with disabilities, purported weaknesses, the dreaded 'comorbidities'. Bodies were now sites of infection, of potential lethal danger; they were the loci of vulnerability, unpredictability. At the same time, virtual meeting platforms made it possible for those considered 'vulnerable' and differently abled to contribute on a more equitable basis, to become 'visible' as we were all rendered talking heads on screens.

The genesis of this book spanned those fearful first years of the COVID-19 pandemic, a time when preoccupations with health and illness were dragged into the foreground of all our lives. On top of the everyday human ubiquity of sickness, loss, death and mourning, we were overwhelmed by layers of extra, and often impossible, grief and suffering. Nearly everyone, worldwide, lost people they loved, often under conditions of strict quarantine that forbade hospital visits, bedside vigils or funerals; many underwent lengthy ordeals in hospitals; yet more were disabled or left chronically ill – with systems of medicine and health care (often already compromised and broken) that had no experience, no history, no narratives for this disease. Terrified by what was little understood, some seized upon conspiracy theories, taking anti-vaccine stances, peddling quack remedies. It is worth remembering that, at present, there is not yet one doctor alive who was able to study COVID-19 in medical school. So the third word in the subtitle of this book springs into leering focus: we were (and still are) all traumatised.

We have developed varying degrees of amnesia about the stress of those early months and years, before the arrival of vaccines and the mutations of the virus itself into forms that were unpleasant and disabling rather than lethal. Globally, our reluctance to count the cost of the pandemic, our refusal to acknowledge its continuing presence, like Banquo's ghost at our table, is indicative of our shared trauma.

This is why *Notes From the Body* is such an important book. It is about all of these conundrums, heightened by a global medical crisis – but the pieces in it tell their truths slant. This is *not* a book about COVID, but it is about the frailty and also the extraordinary adaptive qualities of our fragile bodies. Joanne Bloch writes about the multiple ways she and others learn to reread context, clothing, activities and communities, in a deliberately vivid piece on loss of sight. Warren Swinney presents a series of observations and snapshots (literally) on the musician Gary Herselman, who experienced cynanthropy (the belief that he was a dog) at times of extreme stress. Kobus Moolman's poem-epic about the ancient myth of the selkie plunges us into what it is to inhabit a broken body, or indeed the "wrong" body, and our obsession with "fixing" (double meaning deliberate) others within the limits of what we believe their outward bodies should signify. Antjie Krog, in the poem "it's when everything seems to be falling apart", encompasses both COVID and its all-too-human context in lines like "we feel how bush virus and human hate/scorch the hair on our arms".

It is a book about how childhood sexual abuse leaves psychic shards of glass in our souls, usually in perpetuity (and thus about what happens when a violated child is disbelieved – the extreme and toxic version of having our physical symptoms dismissed as "neurotic", or our emotional wounds masked by inappropriate medication). Phillipa Yaa de Villiers's poetic meditation on her history of sexual violence flows effortlessly into other contested areas of her identity; Gaireyah Fredericks howls with rage that allows for no excuses, no forgiveness in her account of how her father raped her while her mother turned a blind eye; Hélène Smit's fragments of memoir reflect the disassociation that is often the price paid for sexual violence in childhood.

The book also disturbs the binaries of our physicality: mind-body, sickness-health – creating and focusing on the unexpected spaces of rage, enquiry and creativity that billow out between those delineations.

Many of these contributions are painful to read, some even demanding. There is no resolution, no redemptive arc, no hope in Diana Bloem's prose poem on her bipolar son, or the boiling rage in Fredericks' "Lights Out!" There's an awful lot of shit going down (once again, literally: all the metaphors of the body double up as the real deal in this collection). Some pieces are fascinating, some funny; there's a memorable mental image of Duncan Brown wearing "Victoria's Secret" hospital bloomers as an ambulance carts him back and forth between two towns.

I first read this book months ago, and certain essays reached deeper into me than others. On re-reading for purposes of writing this review, different essays now poke at tender spots, elicit electric shocks of recognition. My responses were more than dialogic: they were (and are) autobiographical. Many readers will have the same reaction, finding their own unique and often private stories surfacing: they will *recognise* themselves. In this respect, this collection is invaluable, a kind of processing – even therapy – made possible by words on the page.

My life since March 2020 has been profoundly shaped by "health, illness and trauma"; I contracted COVID several days before South Africa's lockdown began, followed by seven weeks of pneumonia, followed again by the dubious distinction of being one of the first South Africans to be formally diagnosed with Long COVID – a condition unheard of and unrecognised in those early days of panic and mystification. There was a terrible loneliness in having one's life invaded by a condition that did not yet officially exist. As I continue to live with Long COVID (which I call LoCo,

in acknowledgement of its maddening and mystifying manifestations) more than four years after my first infection, it is the lens through which I took in most of the material in *Notes From the Body*.

This autobiographical identification did not stop with my own illness. In mid-December 2020, with vaccines for most South Africans still months away, my sister was admitted to Kingsbury Hospital with COVID. She collapsed in their emergency unit, was on a ventilator for a month, in an induced coma for seven weeks, in first critical and then serious condition in ICU for over two months. Her X-ray history shows that, at one point, she had the use of only five per cent of one lung. The sepsis that followed her COVID pneumonia meant that the dread in which we lived was unrelieved until only a few days before she was discharged (after 65 days – the hospital's record for the longest period a patient had spent in ICU and survived). She is now viewed by the medical community as a mystery, a bona fide miracle: we had been advised to brace ourselves for death or severe brain damage. Yet today she is less impaired than I am.

Samuel Njenga's account of (barely) surviving his spell in ICU with COVID hurled me back into that world. I recognised much, but was also struck by the differences between his experience and my sister's: he was conscious for some of his ordeal, and able to communicate with his family via iPad. He notes that after these one-sided conversations (speech is not possible after the tracheotomy that opens the throat to the blessed invasive ventilator), his oxygen saturation levels rose – further underlining the tension between the loneliness of dread disease and the need to locate our ailing selves within a community, to experience love and care when in extremis.

Likewise, Nkosinathi Sithole's account of how his life was saved both by the treatment he received for TB and the intercessions of his ancestors took me back to the months of my sister's coma, in which I fell again and again into a safety net of the prayers of others; performed visualisations that included sending four dogs on patrol through my sister's veins and organs; became convinced that the heron I saw in a nearby pond each morning was a protective spirit. The daughter of a scientist, I actively sought any 'non-rational' sign that offered hope, even as my sister's doctors 'cautioned' us against hope.

As a scholar, a child of deconstruction and, in my youth, a student of Derrida and Marx, I was accustomed to tracing the violence of either/or language or the dialectic within texts; decades later, it is clear that apparently oppositional symbols can collaborate, inflect, amplify or erode each other. Sithole's presentation of the multivocal responses to his illness, diagnosis and treatment shows how binarisms – such as 'Western' and 'traditional', 'education' and 'ignorance', 'reason' and 'superstition', even 'mind' and 'body' – are crude and ultimately unhelpful in our struggles to give meaning to suffering.

Margie Orford, whose intense and moving Foreword is worth the price of admission alone, also notices this: "Grief, like illness, like mental and physical pain, undoes the Cartesian divide between mind […] and the unruly partisan feeling body." This is no theoretical response, but like mine, an autobiographical one; reading the book, she relives the pain of the sudden and recent death of a beloved sister. But she also points out the potential not so much for healing as *connection* that these pieces make possible: "Pain resists relation and connectedness. The person in pain is isolated from others […] One is cast out by pain, cast aside. The power of language, like a loving touch, is that it creates connection. It is through writing (and reading) that we are able to see through the eyes of others, to feel with them."

Voices speak to one another, too: in an essay comprising emails and poem fragments between Andrea Garman and Gillian Rennie, the authors speak with sometimes disturbing honesty about the former's breast cancer and mastectomy, and how this affects their friendship, the dyad they construct for themselves. Their dual voices, laid out in parallel on the pages, are a riposte to a medical fraternity and society that is bothered by "asymmetrical" bodies.

The socio-political landscapes in which our bodies live are also presented, although not in the form of conventional political commentary. Vonani Bila's contribution – a riveting account of the trauma he experiences when thugs shoot him multiple times, and his long and incomplete journey towards recovery – is one of the pieces that holds the state accountable for the parlous treatment of its citizens' bodies. His account of the collapsing health services and other vital forms of infrastructure – such as navigable roads – in rural Limpopo where he lives, is scathing, but also a creative tour de force: like many of the pieces here that testify to bleak realities, the "salvage" comes in the form of the triumphant agency of the artist.

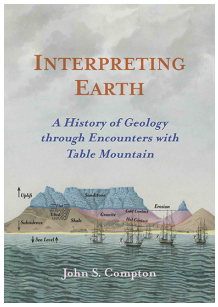
Nearly every piece combines sometimes shocking honesty with nuance at a historic moment when public discourse – especially about bodies, trauma and the role of science – has been polarised into wild rhetorical claims and boastful ignorance. These essays supply a refreshing push-back against the crassness of contemporary conversations on bodies and health.

The criticisms: as with all essay collections, some contributions shine more brightly than others – but this is as much about the reader as the writer. Everyone will have their preferences, but no one will be left unmoved. The reach and format of the work is ambitious: the contributors have been given glorious licence in terms of genre, their writing drawn from a variety of poetic forms, artworks, email correspondence, diary entries, fragments of memoir, and more formal essays. Some of the authors negotiate this freedom to play and experiment with a panache that reflects their experience and skill: it is hardly surprising that creative artists such as Phillipa Yaa de Villiers and Vonani Bila provide such bravura combinations of poetry, reflection, prose and memoir.

Possibly the most forceful critique I might make is that not enough scope has been given to discuss the astonishing innate abilities of the body, especially in terms of learning and storing skills. Bronwyn Law-Viljoen is the only contributor who specifically addresses this, recounting her history in terms of physical skills mastered – netball, bicycle racing, boxing, dancing – speaking of the "kinetic beauty" in which "[t]he athlete repeats and repeats an action, until action […] seems to be thought itself". Anyone who has watched a ballet dancer en pointe, listened to the cadenza of a violin concerto, seen a carpet woven in an Agra workshop, witnessed Brian Lara in his heyday play a hook shot, or Serena Williams blaze through a backhand volley will know that moment of catching one's breath at (in the words of the Jesuit poet Gerard Manley Hopkins): "the achieve, the mastery of the thing" – in these cases, the mortal body.

I have remarked on the absence of an historical record in dealing with COVID – it has been part of human existence for only four years and several months as I write this. This gap – not even an erasure as much as a blank – made me realise how profoundly we need shared narratives of illness and physical trauma: and this is partly why this collection provides a welcome and necessary gap, not just to academics, healthcare practitioners and those who read and write poetry and other creative works – but to all of us who inhabit frail, morbid, mortal and miraculously adept bodies: our lifelong homes.

BOOK TITLE:
Interpreting Earth: A history of geology through encounters with Table Mountain



AUTHOR:
John S. Compton

ISBN:
9780796112521 (hardcover, 335 pp; ZAR450)
9780796128317 (softcover, 335 pp; USD28)
9780796112538 (ebook, 335 pp; USD10)

PUBLISHER:
Earthspun Books, Cape Town

PUBLISHED:
2023

REVIEWER:
Saul Dubow[1]

AFFILIATION:
[1]Smuts Professor of Commonwealth History, University of Cambridge, Cambridge, England

EMAIL:
shd28@cam.ac.uk

HOW TO CITE:
Dubow S. Table Mountain: Reading the rocks on an unimaginable scale. S Afr J Sci. 2024;120(11/12), Art. #19079. https://doi.org/10.17159/sajs.2024/19079

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
04 December 2024

Table Mountain: Reading the rocks on an unimaginable scale

John S. Compton's geological biography of Table Mountain delights and excites from start to finish. Beginning with Khoisan cosmology as well as Portuguese and Dutch encounters with the Mountain, the author traverses widely as he takes on enlightenment thinking about stratigraphy, questions about the shape of the earth (spherical or oblate), early debates in geology focusing on the unfathomable age of the earth, the question of Neptunism vs Plutonism, the puzzle of rising and falling sea levels, and more recent controversies around continental drift and plate tectonics. Compton is as sure a guide as one might hope for. Table Mountain is, in his capable hands, both a subject of analysis and a platform for the interpretation of broader geological theory and history.

Dante Alighieri's *Divine Comedy*, we are reminded, imagined a terraced mountain in the south, its lowest level comprising Hell, rising successively through Purgatory and culminating in a flat-topped earthly Paradise with Empyrean Heaven above. Although Dante could not have had any direct knowledge of Table Mountain, European travellers versed in his poetry were primed to see Table Mountain according to this schema. And yet, for all its familiarity, the Mountain remained elusive.

It is not at all certain that Bartolomeu Dias glimpsed Table Mountain on his journey around the Cape, or even on his return voyage. Antonio de Saldanha was likely the first European, perhaps the first person, to have scaled its summit (in 1504). While there is rich archaeological evidence of Middle Stone Age life on the lower reaches of the Mountain, there is no record of human occupation in caves higher up. Despite its commanding presence, the absence (or cursory descriptions) of Table Mountain in travel accounts is equally salient. Only in the period of Dutch occupation do we find regular reports of mountain ascents: it was a challenging, sometimes dangerous climb that, from the early period of English occupation, became a routine hike for visiting sailors and, later, a fashionable excursion for turn of the century society figures like Lady Anne Barnard. The domestication of Table Mountain was part of the process whereby the Cape of Storms transmuted into the Fairest Cape, a sublime rather than a fearsome presence.

It is well known that artists failed to render Table Mountain accurately. Robert Sayer (1754) entirely missed the horizontal layers of sandstone that define the massif. William Hodges did better in 1772. It took longer still to make sense of the layered composition of the Mountain, from Malmesbury shale at the bottom, through granite and muddy sandstone, to the ordered layers of sandstone beds further up. The geological significance of these discrete rock layers, brilliantly sketched by Albrecht Herport in 1669, was thus apt to be missed.

Only in the 19th century did Table Mountain come to serve as a prism through which understanding of animated academic discussions could be conducted: for example, whether catastrophic volcanism or incremental deposition shaped by the sea and by wave action was the operative creative force. The telling intrusion of granite into shale, evidenced in Platteklip Gorge (the most accessible route up the Mountain) was missed by travellers and observers until the arrival of Basil Hall in 1812. Soon after, another visitor bound for China, Clarke Abel, identified a striking shale-granite geological contact at Sea Point. He viewed the 'intimate intermingling' of rocks as supportive of Plutonist views of hot magma intrusion, although Abel also saw the slow action of Neptunism in the topography of overlying sandstone. Table Mountain thus became a case study for European, specifically Scottish, enlightenment argumentation about the earth's formation.

The absence of fossils embedded in Table Mountain contrasts with their abundance in regions in the interior where powerful evidence of glaciation and of contortion and deformation is spectacularly demonstrated throughout the mountains of the Cape Fold Belt. For Alex du Toit (1878–1948), surely one of South Africa's greatest scientists, this offered strong reason to pursue the iconoclastic theory of southern hemispheric continental drift. Du Toit's broad imagination linked the folded mountains of the Cape, through Table Mountain and the Falkland Islands, to Latin America. Behold! Port Stanley adjacent to Port Elizabeth!

Du Toit's reworking of Wegener's theory of continental drift in *Our Wandering Continents*[1] centred on the idea that 'Africa forms the Key', an idea pursued by Jan Smuts in association with Raymond Dart's claims that *Australopithecus africanus* constituted the key missing link in hominin evolution. Decisive evidence required to settle the contentious debate about the mechanics of continental drift was only forthcoming with the emergence of plate tectonics theory in the 1960s – a paradigmatic shift in geological thinking that Compton attends to in the final chapter of the book.

Compton's synoptic eye draws attention to important geologists at the Cape as well as the importance of geological awareness. His reading of John Barrow, whose travel writings and descriptions are well known to historians, demonstrates just how foundational geology was to the thinking of this significant imperial scientist and policymaker. It was Barrow who determined that Devil's Peak and Lion's Head originally formed one mass with the central Table. Compton is illuminating, too, about the geological ruminations of other well-known visitors and travellers better known as botanists and naturalists, such as William Burchell, Hinrich Lichtenstein and François Le Vaillant. Just as Table Mountain's geological features hide in plain sight, so their views on geology have eluded the attention of readers more attuned to descriptions of fauna and flora.

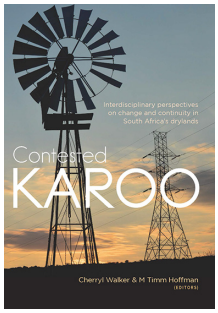
Compton's focus moves between aesthetic evocations of the Mountain to history and earth science. He delights in walking his readers up Platteklip Gorge, for long the preferred route for mountaineers and hikers, as well as an instructive petrological text for geologists. He also uses Table Mountain as a means to engage with broader geological theories that do not necessarily pertain strictly to the Mountain itself. For uninitiated readers like myself, this is not a matter for complaint. Compton's style is breathy and bracing, his enthusiasm is infectious, and his

core knowledge, if not all his interpretations, unimpeachable. As popular science, this is an exemplary contribution, comparable to Hugh Eales' *Riddles in Stone*[2] which also tackles South African geology. My only complaint relates to Compton's idiosyncratic and minimalist referencing system, which makes sources difficult to check. This niggle is, however, outweighed by the book's abundant and well-chosen illustrations.

References

1. Du Toit AL. Our wandering continents: An hypothesis of continental drifting. London: Oliver & Boyd; 1937.
2. Eales HV. Riddles in stone: Controversies, theories and myths about southern Africa's geological past. Johannesburg: Wits University Press; 2007.

BOOK TITLE:
Contested Karoo: Interdisciplinary perspectives on change and continuity in South Africa's drylands



EDITORS:
Cherryl Walker
M. Timm Hoffman

ISBN:
9781991450005 (paperback, 344 pp; ZAR200)
9781991450012 (open-access ebook, 344 pp)

PUBLISHER:
UCT Press, Cape Town, South Africa

PUBLISHED:
2024

REVIEWER:
Emma Archer[1]

AFFILIATION:
[1]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa

EMAIL:
Emma.Archer@up.ac.za

HOW TO CITE:
Archer E. The country ahead of us, the country behind: A review of 'Contested Karoo'. S Afr J Sci. 2024;120(11/12), Art. #19150. https://doi.org/10.17159/sajs.2024/19150

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
04 December 2024

The country ahead of us, the country behind: A review of 'Contested Karoo'

Contested Karoo arrives as a valuable and timely contribution to the field of work focusing on the Karoo – a field that has tended to be dominated by studies of the physical environment. The book constitutes a collaboration between scientists working more on the physical environmental side and social scientists, providing critical perspectives both on the Karoo past and also, importantly, on the different plausible futures facing the Karoo, given ongoing social and environmental challenges. The book effectively provides a type of 'road ahead, road behind' perspective on a rapidly evolving landscape – biologically and socio-economically.

The Karoo has often tended to be viewed from the outside as empty – a repository for a range of projects, programmes and imaginaries. Yet, of course, people have lived there for thousands of years – and the tendency to consider the Karoo as a type of empty stage for favourite narratives (for example, around what constitutes 'good' rural development), has had significant consequences. As a result, while a rich and rewarding volume, the book does not provide comfortable bedtime reading. It is clear, across the range of sites and examples considered in the book, that most land uses have overpromised and underdelivered to local communities.

The book argues convincingly that the Karoo should be regarded as a type of "reconfigured resource frontier" (Chapter 12) – where powerful vested interests effectively focus on extraction of value (for example, minerals, and renewable and non-renewable energy). As a result, the notion of comfortable narratives imposed by external view is key – for example, in the case of renewable energy, several chapters emphasise the extent to which renewable energy has been seen in the Karoo (and, of course, elsewhere) as a kind of 'silver bullet' strategic investment – one that can ensure energy security and the transition to low carbon energy generation, but one that can also provide a measure of socio-economic benefit. A number of authors unpack this notion – including, for example, the chapter on renewable energy investments outside De Aar, where the contrast between the renewable energy infrastructure and the significant energy poverty amongst the local communities is both striking and timely, given the prominence of discussions in South Africa around the 'just transition' at present.

The book also considers the past and future of farming in the Karoo, noting that livestock farming is likely to continue to constitute a key part of plausible land use futures for the area – but that it is already changing, and will continue to adapt in response to environmental and socio-economic change. For example, it is clear that the trend of declining commercial stock and the rise of different forms of game farming is a continuing one[1] – but this is another example of the extent to which the book unpacks what might be comfortable narratives around Karoo land use. Frequently, externally imposed storylines regarding the changing nature of Karoo farming are very uncritical and optimistic about the environmental and socio-economic (including labour) benefits of the rise of different typologies of game farming. We know, of course, that game farming activities come in a variety of forms, and the critical debate throughout a range of chapters, around not just the current and possible benefits of game farming but also types of eco-tourism activities, provides real practical considerations in terms of how such activities might be planned, implemented and viewed.

What really resonates throughout the book as we consider emerging uncritical storylines is the notion of power. When we unpack, for example, how renewable energy or tourism in the Karoo have tended to be presented somewhat optimistically and uncritically, one of the key approaches taken by a range of chapters is effectively unpacking these by asking 'who has the power'? Power here can include both who has the power to define what success might look like, but also who may define who benefits, and how. If there is one single more important practical message that one can take away from the book, it reminds us that we can never be overconfident and uncritical in considering plans and programmes in fragile landscapes – that we have to ask the difficult and uncomfortable questions.

Effectively, the book constitutes a key contribution beyond the Karoo itself, to understanding of resilient drylands, with clear implications at both national and global levels. A primary example here would be the aforementioned concerns regarding priority investments in marginal environments (e.g. mining, renewable energy and the SKA), and their ability to provide equitable benefits to all, including so-called 'surplus' communities, while ensuring environmental sustainability. This cluster of chapters alone has significant implications more widely – far beyond the academic community, and, in fact, far beyond the Karoo.

The visual material in the book provides a rich complement to discussions in text – my one comment was that I would very much have liked higher-resolution maps to consider, but this was likely a feature of the electronic review copy. The photographs are beautiful, and emphasise the powerful love and commitment of many of the authors both to the landscape, and to those living within it.

In terms of audience, the book is certainly targeted for specialists and peers, with robust and useful material around specialist areas. But I think the use of the book goes beyond domain specialists – which is important. I think (and hope) that it will also be of significant interest to the civil society organisations active in this area, including those concerned with social and environmental justice. There are also useful notes regarding changing paradigms in conservation and agriculture – these go beyond interest only to environmental and conservation scientists, and are likely to be of clear interest to farmers, civil society and conservation practitioners. I would argue that this is a highly desirable outcome.

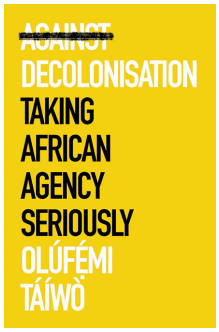
Acknowledgement

Sincere thanks are due to David Guterson for his permission to make use of the title of his 1989 book of short stories in the title of this review.

Reference

1. Hoffman MT, Skowno A, Bell W, Mashele S. Long-term changes in land use, land cover and vegetation in the Karoo drylands of South Africa: Implications for degradation monitoring. Afr J Range Forage Sci. 2018;35(3/4):209–221. https://doi.org/10.2989/10220119.2018.1516237

BOOK TITLE:
Against decolonisation: Taking African agency seriously



AUTHOR:
Olúfẹ́mi Táíwò

ISBN:
97817873869621 (paperback, 270 pp)

PUBLISHER:
Hurst Publishers, London; GBP15

PUBLISHED:
2022

REVIEWER:
Veeran Naicker[1]

AFFILIATION:
[1]Centre for the Study of the Afterlife of Violence and the Reparative Quest (AVReQ), Stellenbosch University, Stellenbosch, South Africa

EMAIL:
vnaicker@sun.ac.za

HOW TO CITE:
Naicker V. Modernity and its discontents: Tradition and the problem of liberation in postcolonial Africa. S Afr J Sci. 2024;120(11/12), Art. #19258. https://doi.org/10.17159/sajs.2024/19258

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
04 December 2024

Modernity and its discontents: Tradition and the problem of liberation in postcolonial Africa

At the turn of the millennium, the Marxist historian and cultural critic Arif Dirlik surmised that if there was once scholarly consensus in the 1970s over the meaning of colonialism, as the political and economic control of the territory of one society by another, the intrusion of postcolonial discourse into the academy in the 1980s signalled the elevation of colonialism to a "totalizing structure", which came to be synonymous with a metaphysical adumbration of modern society as colonial Eurocentrism. He cautioned that this theorisation of colonialism tended to de-historicise the historical dynamics conditioning the emergence of specific colonial systems of governance (each producing complex, differing effects) with the radical representation of a unified metaphysical system that erased all analytical specificity. His diagnosis was that this development in intellectual history was symptomatic of the morality of a global elite whose knowledge production is defined by a lack of interest in critiquing the political system they constitute, rather asserting their authority with different iterations of "cultural nationalism".[1]

Although Dirlik's critique is representative of an orthodox Marxist position that has reductively read postcolonialism's claim that colonial racism is irreducible to capitalist exploitation as bourgeois idealism, the emergent institutional dominance of Latin American Decoloniality theory as the global successor of postcolonialism has led credence to his symptomatic reading of the new *postcolonial*, or rather *decolonial* intellectual. Ramon Grosfoguel and Walter Mignolo have been influential in arguing that it is less valuable to be a political activist in the periphery campaigning for modern human rights and freedoms, i.e. reproducing colonial epistemologies, than to be in the metropole practising "epistemic disobedience", subverting the epistemic conditions that structure our global capitalist world system. They propose that true decolonisation from the colonial matrix of power, knowledge and being, that is, colonisation as a metaphysical system, necessitates dismantling the dominant Western knowledge system at the metropolitan core and centring the hitherto marginalised and excluded epistemic contributions of non-Western indigenous knowledge systems, which offer alternative ways of understanding and being in the world.[2,3]

In contrast to this trend, the exemplary scholarship of Professor Olúfẹ́mi Táíwò, in his new book, *Against Decolonisation: Taking African Agency Seriously*, which he describes as a footnote to his earlier, now classic text, *How Colonialism Preempted Modernity* (2009), is an exception that speaks profoundly to the challenge laid out by Dirlik. One advantage of Táíwò's account over his opponents is his rigorous theoretical and conceptual delimitation of what decolonisation entails. He differentiates between two distinct forms, Decolonisation1 and Decolonisation2 (p.1–12). Decolonisation1 involves the positive transformation of a colony into a self-governing entity that directs its own political and economic existence. The first chapter of the book illustrates the positive conception of Decolonisation1 as reflected in the work of prominent anti-colonial intellectuals and revolutionaries, Frantz Fanon, Amilcar Cabral and Kwame Nkrumah. Táíwò criticises what he views as a farcical application of Decolonisation2, the result of conflating modernity and colonialism. This form demands that an ex-colony reject entirely any cultural, political, intellectual or linguistic artefact idea, process, institution or practice that retains even the slightest trace of the colonial past, under the threat of remaining eternally colonised. Táíwò associates the genealogy of Decolonisation2 on the African continent post-independence, not with the prominent Latin American Decoloniality school, which has been disseminated through the work of Sabelo Ndlovu-Gatsheni and the African Decolonial Research Network, but with the thought of Ghanian-born philosopher Kwasi Wiredu and the Kenyan born literary critic, Ngugi Wa Thiong'o.[4]

My intention in this review is not to probe whether Táíwò's historical sourcing of Decolonisation2 in the postcolonial African context is accurate. However, it is worth cautioning the reader that, while the contributions of Táíwò's criticisms are profound, their work (particularly Wiredu) is of marginal importance to contemporary debates concerning the decolonisation of postcolonial African society. My interest is in examining the presuppositions for this interpretation and why Táíwò does not engage directly with the Decoloniality school in the body of his text. Paradoxically, this choice stems from a shared Fanonian agreement and disagreement with Latin American decolonial scholars on the repressive nature of the postcolonial bourgeoisie, outlined in Fanon's chapter on *The Pitfalls of National Consciousness* in *The Wretched of the Earth*. The Decoloniality school argue that liberation from colonisation was incomplete insofar as the national bourgeois adopted the epistemic values of coloniality after the formal demise of colonisation. The core of Táíwò's argument follows a similar premise, as he argues that once independence from European colonialism was obtained legally in Africa, Africa's rulers

> *proceeded to put in place numerous political contraptions – all designed to subvert and deny the freedom of their own people – to turn their citizens into subjects and to substitute their own wills for those of their people when it came to the installation of governments all across the continent… In other words, the promise of independence was never redeemed for ordinary Africans at the micro-level of their quotidian lives.* (p.194)

Táíwò's normative focus is realising the potential of modernity for ordinary postcolonial Africans, which he defines as the second struggle for freedom following independence. His work engages with a tradition of African political philosophy concerned with the problem of underdevelopment and the failure of democratic institutions on the continent. Walter Rodney's field defining intervention challenged prevailing views that failure was predicated on African cultural values, rather claiming that Europe's colonial system of racial capitalism, and blatant disregard for African sovereignty were to blame. Táíwò's original contribution to this debate is the assertion that the overwhelming Fanonian characterisation of colonialism as direct rule by contemporary decolonial scholars neglects the way local traditions were appropriated within the exploitative and oppressive structures of colonial governance. He argues that, by simplifying colonialism in this manner, scholars have fundamentally neglected the

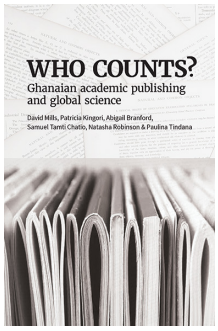
sociopolitical impact of British indirect rule in Africa, a problem which he expounds upon in detail in *How Colonialism Pre-empted Modernity*. According to Táíwò, British indirect rule inculcated a social formation he refers to as sociocryonics. Sociocryonics is the preservation of traditional African cultural practices as historically immutable, which has led to the uncritical perpetuation of indigenous systems of governance after colonialism. Táíwò analyses the maintenance of chiefly rule and child marriage by postcolonial African leaders, which has stymied the dynamic growth and change of cultural practices across the continent. His core argument is that sociocryonics serves the venal interests of postcolonial African ruling classes by politically repressing the rights, liberties and formal equality that characterises democratic citizenship. He argues that post-independence African rulers have adopted these colonial tactics, legitimising oppressive and exploitative 'pre-colonial practices' by referencing their 'Africanity'.

The challenge that Táíwò's book bequeaths to students of contemporary decolonial theory is a philosophical problematisation of the role traditional, indigenous culture plays in contemporary society as inherently subversive, progressive and liberatory. Responding to the prevailing wisdom of an entire generation, Táíwò's book raises a more unsettling problematic: are scholars associated with the progressive left propagating traditionalist discourses that would not liberate Africans, but lead them further into oppression? While he does not flesh out the implications, I will conclude by noting that, after the breakdown of formal colonial relations, Pan African principles were institutionalised through the founding of the Organization of African Union (OAU) in 1963. With the demise of the Cold War, the shift from the OAU to the African Union (AU) was influenced by the political discourse of the late Muammar Gaddafi (1942–2011), and the emergence of post-apartheid South Africa's former president, Thabo Mbeki's (1942–) emphasis on a postcolonial African renaissance. The AU states that it promotes an identity that is liberal, democratic and inclusive, but similarly its reconstructive, governmental strategy continues to endorse an essentialist conception of African unity as grounded in tradition, or "indigenous African culture".[5]

References

1. Dirlik A. Rethinking colonialism: Globalization, postcolonialism and the nation. Interventions. 2002;4(3):429–448. https://doi.org/10.1080/1369801022000013833
2. Grosfoguel R. Decolonizing Western universalisms: Decolonial pluriversalism from Aimé Césaire to the Zapatistas. Transmodernity. 2012;1(3):88–104. https://doi.org/10.5070/T413012884
3. Mignolo W. Epistemic disobedience, independent thought and decolonial freedom. Theory Cult Soc. 2010;26(2–8):159–181. https://doi.org/10.1177/0263276409349275
4. Pillay S. The problem of colonialism: Assimilation, difference, and decolonial theory in Africa. Crit Times. 2021;4(3):389–416. https://doi.org/10.1215/26410478-9355201
5. Chipato F. The global politics of African identity: Pan-Africanism and the challenge of Afropolitanism. Glob Stud Q. 2023;3(3):1–11. https://doi.org/10.1093/isagsq/ksad046

BOOK TITLE:
Who counts? Ghanaian academic publishing and global science



AUTHORS:
David Mills, Patricia Kingori, Abigail Branford, Samuel T. Chatio, Natasha Robinson, Paulina Tindana

ISBN:
9781928502647 (open access ebook, 239 pp)

PUBLISHER:
African Minds, Cape Town

PUBLISHED:
2023

REVIEWER:
Lily Kpobi[1]

AFFILIATION:
[1]Regional Institute for Population Studies, University of Ghana, Accra, Ghana

EMAIL:
Lkpobi@ug.edu.gh

HOW TO CITE:
Kpobi L. Beyond 'publish or perish': Unpacking the complexities of academic publishing in Africa. S Afr J Sci. 2024;120(11/12), Art. #18970. https://doi.org/10.17159/sajs.2024/18970

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
04 December 2024

Beyond 'publish or perish': Unpacking the complexities of academic publishing in Africa

In the book, *Who counts? Ghanaian Academic Publishing and Global Science*, David Mills and colleagues thoughtfully examine and critique the African academic publishing landscape, drawing on the experiences of academics at two Ghanaian public universities and from journal editors in Africa more broadly. Based on the findings of their qualitative study, they interrogate the slow but growing area of scholarly writing on the continent by first discussing the historical precedents of academic publishing on the continent and then looking at current perspectives and experiences. By examining local and regional publishing cultures and the challenges and opportunities that exist for African academics seeking to contribute to global science discourse, the authors delve into the complex and multi-layered factors which drive the African scholarly writing agenda.

The book's 10 chapters provide critical discussions on issues that are familiar to many budding academics on the continent, from the ever-popular 'publish or perish' mantra which has arguably resulted in an emphasis on quantity over quality of scholarship in many universities, to the spectre of predatory publishing and the influence of the global bibliometric economy and how this has potentially undermined the growth of the African publishing agenda.

In Chapter 2, the authors offer a critical perspective on the historical context of academic publishing on the continent, from the rapid growth of African journals in the immediate postcolonial era, to the struggle for credibility by emerging African journals, and the impact of the political and financial turbulence of the late 20th century on universities and research more generally. They also discuss the emergence of stricter criteria for ranking universities and the increasing use of indexing metrics which have resulted in an emphasis on broadening the quantity of visible research outputs over quality and contextual relevance. The chapter also provides a striking perspective on the utility as well as the consequences of the burgeoning open-access model of publishing, showing how this model can provide increased visibility to research writings of Ghanaian (and African) authors, but also remains largely inaccessible for most researchers from the Global South due to the high article processing fees. As the authors critically explain, this model reinforces colonialised systems of exclusion of less-known voices in discourse about Africa.

Chapters 3, 4 and 6 examine the study participants' perspectives and experiences of why it is important for Ghanaian researchers to be visible in the publishing space, and the regard placed on such visibility in international journals which are considered credible or reputable. This was described as necessary for career progression despite the very real constraints experienced in university systems which do not provide much support but require results. This situation sometimes makes authors easy 'prey' for so-called predatory journals. The authors offer a thought-provoking analysis of the impact of reforms and policies which have transformed the way the (global) academy operates. These changes have been reported by other scholars to have had significant impacts on lesser-resourced institutions or contexts as well as on the work of individual academics, especially female academics. Within an increasingly globalising research landscape, scholars in critical higher education science have long argued that the shift towards a business model for higher education can result in greater precarity of careers, fostering unhealthy competitions and rivalries.[1] This was clearly seen in the narratives of the book's participants.

The second half of the book looks at the editorial side of publishing by sharing the experiences of Ghanaian and African journal editors, including their views on sustaining their journals within a competitive scholarly economy. In Chapter 7, editors of various Ghanaian journals share their struggles for legitimacy, funding or financial support, and the difficulties of maintaining standards when most of the required work is conducted on a voluntary basis. The book also highlights the support that platforms like African Journals Online (AJOL) provide to local journals, by facilitating the establishment of publishing rigour and building editorial capacity. As Mirowski[2] argues, with the dominance of international publishing houses that are increasingly taking over regional publishing opportunities, smaller local journals have to prove their relevance, and often have low impact factors, but collectives such as AJOL and AOSIS can help to improve this.

The two concluding paragraphs of *Who counts?* provide an optimistic view of the future for Ghanaian (and African) publishing culture if proactive and sustained measures are taken, such as diversifying journal metrics systems and university rankings to include citation indexes that have been developed by and for African journals. As the authors indicate, "The future of Africa's research ecosystems depends on strong and well-funded national research systems, bibliodiversity, multilingual publishing, and alternative circuits of academic credibility." (p.201)

The authors also emphasise the importance of Ghanaian universities moving away from the reliance on metrics and impact factors for career advancement. This was anticipated to reduce the need to 'publish at all cost'. Essentially, the book's participants advocated for an evolution of the research and publishing cultures of African universities and for a recognition of local contextual determinants in the global research economy.

The core message of the book then, is that the requirement for African academics to be more visible in the publishing space is not one that will likely disappear. However, this needs targeted transformation in order to emphasise the voices and experiences of African scholars in a manner that enhances the contribution to and development of global science by Africans for Africans. As the authors indicate, the prestige of both African authors and African journals is rooted in "postcolonial history, geopolitics, and an increasingly transnational research agenda" (p.8), and the increasing digitisation and commercialisation of academic publishing has

further deepened North-South inequalities. And yet, as this book shows, African researchers are navigating this space in an effort to survive and even thrive in spite of the inequalities. Ultimately, a thriving global research economy must include the active and equitable participation of voices from Africa, but this can only be achieved when the publishing landscape in African universities evolves to reflect the reality on the ground.

References

1. Ivancheva M, Lynch K, Keating K. Precarity, gender and care in the neoliberal academy. Gend Work Organ. 2019;26:448–462. https://doi.org/10.1111/gwao.12350
2. Mirowski P. The future(s) of open science. Soc Stud Sci. 2018;48:171–203. https://doi.org/10.1177/0306312718772086

AUTHOR:
David J.N. Limebeer[1]

AFFILIATION:
[1]School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
David Limebeer

EMAIL:
David.limebeer@wits.ac.za

HOW TO CITE:
Limebeer DJN. A tale of three Davids. S Afr J Sci. 2024;120(11/12), Art. #19010. https://doi.org/10.17159/sajs.2024/19010

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
David Mayne, control theory, systems engineering, historical perspective, University of the Witwatersrand

PUBLISHED:
04 December 2024

A tale of three Davids

Significance:

On Monday 27 May 2024, the outstanding South African control theorist and systems engineer David Quinn Mayne left us – he was 94 years of age, and a graduate of the University of the Witwatersrand. We celebrate his life and accomplishments, as well as links to two of his academic descendants – both of whom are South African engineers, both of whom are Wits graduates, and both go by the name of 'David'.

In late May 2024, the international engineering community was saddened to hear of the passing of Professor David Quinn Mayne. While the writing of this article was triggered by David's recent death, my intention is not to write an obituary – several obituaries have been written.[1] What I aim to do instead, is celebrate the life of a famous South African engineer, together with links to two of his academic descendants, both of whom are South African engineers that go by the name of 'David'.

David Mayne (David M) was born in Germiston, South Africa, in 1930. He received BSc and MSc degrees in electrical engineering from the University of the Witwatersrand. In 1954, then aged 24, he left South Africa to work as an electrical engineer at the British Thomson Houston Company in England. In 1957 he was offered an academic position at Imperial College London, where he earned both his PhD and DSc degrees and subsequently attained the rank of 'professor'. He also held visiting positions at Harvard (1971) and was a professor in the Department of Electrical and Computer Engineering at the University of California, Davis between 1989 and 1996. He remained associated with Imperial College as an emeritus professor until his retirement.

Soon after joining Imperial College, Pontryagin's Maximum Principle and Bellman's dynamic programming made their appearance. This was accompanied by the popularisation of state-space theory, the widespread availability of computers, and the advent of linear-quadratic optimal control. In an academic sense, David M was born in an interesting time. Modern control theory was still in its infancy and David M found himself in the middle of an intellectual tussle between the British and American control systems research communities. The American school, headed up primarily by researchers at MIT, believed that state-space theory and optimal control were the answer to everything. This enthusiasm was received with circumspection on the other side of the pond, because, while mathematically elegant, these methods often did not work on real engineering problems. David M, probably wisely, had a foot in both camps. On the American side of the equation, he studied estimation, nonlinear filtering, and optimal control, while on the British side he had an interest in generalising classical frequency-response-based control system design methods (for systems with multiple inputs and outputs). As history would have it, both schools turned out to be right, but this was only appreciated after several cross-links were established and fully understood. This is often the way in research.

A little over a decade after David M's birth, David Jacobson (David J) was born in Johannesburg in 1943. He also received a BSc degree in electrical engineering from the University of the Witwatersrand, in 1963, and then, under the supervision of David M, was awarded his PhD degree from Imperial College London in 1967. David M described David J as "the most original thinking PhD candidate of the many I have supervised". One of David M's oft repeated stories is how David J proposed the non-conventional use of strong variations by making large variations in control over short intervals of time rather than the other way around. David M did not initially recognise the value of this suggestion and discouraged David J from pursuing it. But like all good research students, David J recognised the limitations of his supervisor and pressed on with it anyway. This work subsequently culminated in a new and then revolutionary theory known as Differential Dynamic Programming.

On deciding that the conventional quadratic optimal control had become prosaic and overworked, David J introduced the theory of risk-sensitive optimal control. In 2013, Lars Peter Hansen, with Eugene Fama and Robert Shiller, won the Nobel Prize in Economics. In his Nobel lecture, Hansen[2] referred to David J's work on risk-sensitive optimal control and robustness as being relevant to risk analysis in economics[3]. Years later, David J's work on risk-sensitive optimal control was linked to game theory and the then new robust optimal control theory – a topic I know intimately.

David J made numerous contributions to the governance of science and business in South Africa. Between 1975 and 1985 he served at the Council for Scientific and Industrial Research (CSIR) and became its deputy president in 1980. David was also Chair of the South African Mathematical Society, and President of the South African Institute of Electrical Engineers (SAIEE).

Approximately a decade after David J's birth, I was born in Johannesburg in 1952. Like the other two Davids I received a BSc degree in electrical engineering from the University of the Witwatersrand. I then received MSc and PhD degrees from the then University of Natal before moving to the University of Cambridge to study the robust optimal control theory that is closely related to David J's work on risk sensitivity.

In 1985, I met David M for the first time at a conference at the University of California, Berkeley. Following a talk I gave on some of the prevailing ideas, David M, probably after recognising my South African accent, came over and introduced himself. I was immediately struck by his warm and friendly personality. He was already a well-known figure and had every right to be a little 'superior' towards the then young whippersnapper, but he never was. Shortly after our initial meeting in Berkeley, there was an opening for a temporary lecturer in David M's control group at Imperial College. I remember him explaining that they needed somebody who knew about the recent developments in robust control, which he said he found difficult to follow. This was typical of David M's modest demeanour; I knew full well that he could get on top of this material within days if he put his mind to it.

Source: Department of Electrical and Electronic Engineering, Imperial College London

David Quinn Mayne

David M did not appear to be particularly interested in 'managing' the control group, or the people in it. On my first day at Imperial College, he said to me that there was an empty office at the other end of the passage where I could work, and that he would meet me for lunch – this was the sum total of my induction into the Department of Electrical Engineering at Imperial College! This apparent managerial indifference was subsequently inherited by David M's successor Mark Davis, but nothing ever seemed to go terribly wrong. When I eventually took over from Mark Davis, I was by then fully acquainted with the Group's hands-off managerial style. We all did our research and occasionally grumbled about too much teaching work. The group climate might have been intimidating to some, with the conversation frequently punctuated with 'Martingales', 'sequential quadratic programming', 'particle filters', 'non-smooth analysis', Nevanlinna-Pick theory and the 'Black–Scholes equation'. That said, it is hard to imagine a better place to work than in David M's control group.

David M served as the Head of Department at Imperial College from 1984 to 1988. This is another job I inherited from David M and was very proud to do so. At the end of my term, I moved to a professorship in Oxford where I stayed until my official retirement.

It is an interesting fact that David M was well known in the control systems community for decades, but he only became 'famous' after his retirement. Over the entirety of his career, he worked on numerous topics including multivariable control, optimisation, optimal control, adaptive control and nonlinear filtering. He authored over 350 papers and co-authored seminal books on differential dynamic programming and *Model Predictive Control: Theory, Computation, and Design*, which has been cited many thousands of times.[4] In 1997, he began a collaboration with Jim Rawlings (from University of California, Santa Barbara) on a survey of model predictive control. Their work, which was published in 2000, became the most highly cited paper on model predictive control.[5]

Source: Ian Craig

David Mayne receiving the 2014 International Federation of Automatic Control (IFAC) Giorgio Quazza Medal from the then President of IFAC, Professor Ian Craig; Ian Craig is South African and based at the University of Pretoria.

It was recognised in 2011 with the first International Federation of Automatic Control (IFAC) High Impact Paper Award.

David M's long career was littered with accomplishments that attracted a string of high honours. These include the Sir Harold Hartley Medal from the Institute of Measurement and Control; the Heaviside Premium from the Institution of Electrical Engineers that he received twice; the prestigious 2009 IEEE Control Systems Award; the 2014 Giorgio Quazza Medal (for a lifetime of contributions to control systems); and the International Federation of Automatic Control High Impact Paper Award. He was a fellow of the Royal Society, the Royal Academy of Engineering, the IEEE, and IFAC. In 1995, he received the degree of Doctor of Technology (honoris causa) from the University of Lund, Sweden.

David M's dedication to research has left an indelible mark on the academic and engineering communities around the world.

Declarations

I have no competing interests to declare. I have no AI or LLM use to declare.

References

1. Parisini T, Astolfi A. David Quinn Mayne, 1930–2024. IEEE Control Syst Mag. 2024;42(4):142–153.
2. Hansen LP. Nobel lecture: Uncertainty outside and inside economic models. J Polit Econ. 2014;122(5):945–987. https://doi.org/10.1086/678456
3. Jacobson D. Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. IEEE Trans Autom Control. 1973;18(2):124–131. https://doi.org/10.1109/TAC.1973.1100265
4. Rawlings JB, Mayne DQ, Diehl MM. Model predictive control: Theory, computation, and design. 2nd ed. Madison, WI: Nob Hill Publishing; 2020.
5. Mayne DQ, Rawlings JB, Rao CV, Scokaert POM. Constrained model predictive control: Stability and optimality. Automatica. 2000;36(6):789–814. https://doi.org/10.1016/S0005-1098(99)00214-9

AUTHOR:
Anna Taylor[1]

AFFILIATION:
[1]African Climate and Development Initiative, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Anna Taylor

EMAIL:
anna.taylor@uct.ac.za

HOW TO CITE:
Taylor A. Transitioning to climate-resilient development pathways in South Africa: What does it take? S Afr J Sci. 2024;120(11/12), Art. #19891. https://doi.org/10.17159/sajs.2024/19891

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
development pathways, climate change, just transition, resilience, inclusivity

PUBLISHED:
04 December 2024

Transitioning to climate-resilient development pathways in South Africa: What does it take?

Significance:

The notion of climate-resilient development pathways (CRDPs) is gaining traction across science and policy communities as a systemic approach to mainstreaming climate action in the face of changing conditions. This Commentary builds on a recently published review paper reflecting on efforts to move from conceptual development and policy goal setting to operational practices that progress CRDPs. It acknowledges the important convening role played by South Africa's Presidential Climate Commission and calls on the science community to inclusively co-produce the evidence base needed to negotiate and implement CRDPs across scales that unlock a just transition to sustainable well-being for all.

Introduction

The concept of climate-resilient development pathways (CRDPs) is gaining traction internationally, across the science and policy communities, as highlighted in the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.[1,2] CRDPs offer a systemic approach to planning climate action, both adaptation and mitigation, as integral to long-term development by preparing diverse portfolios of actions with flexibility to switch strategies in the face of changing conditions.[3] The question is how to operationally practice such an approach. This Commentary is a reflection of a paper, published in the journal *Current Opinion on Environmental Sustainability*, which reviews how CRDPs are designed and applied in Global South contexts.[4] The review particularly focused on how inequality is dealt with when weighing up priorities and trade-offs, and how climate science is inter-woven with other forms of knowledge in making decisions. The review was undertaken to inform the development of a CRDPs guide, commissioned by South Africa's Presidential Climate Commission (PCC), to support joined-up efforts at realising the ambitious climate and development goals South Africa is striving to attain, in the face of considerable institutional fragmentation and legacies of marginalisation and carbon-intensive industrialisation.

Can the notion of pathways be mobilised across disciplines, policy domains and fields of practice to help understand and act in concerted ways to (rapidly) transition our societies, ecological and technical systems to be more equitable, just and environmentally sustainable? The growing impacts of climate-related events, such as severe flooding in KwaZulu-Natal and multi-year droughts in the Eastern and Western Cape, highlight that our systems are not resilient enough to adequately provide safety and well-being for most people and the infrastructures and ecosystems on which they rely. Working with the idea of climate-resilient development pathways can help draw together diverse actors to negotiate and coordinate actions to set human development (understood as inseparable from earth systems evolution) on a new trajectory, as laid out in the global Sustainable Development Goals (SDGs) and South Africa's National Development Plan.

How CRDPs are being operationalised elsewhere

The question of the utility and applicability of CRDPs is being explored in South Africa through work coordinated by the PCC. Part of that work involved looking at how various pathways approaches have been operationalised elsewhere in the world. Particular attention was given to other Global South contexts, grappling with similar issues to South Africa, associated with high socio-economic inequality, widespread informal land uses, unregulated construction and economic activities, rapid urbanisation, and many people getting by without adequate provision of public services. A systematic review of published academic literature found that many applications of pathways approaches are still in the proof-of-concept phase.[3] This is especially the case in Global South contexts, where relatively small, contained efforts have been made (and reported on in the published literature) to show how pathways thinking can be applied to potentially support decision-making. There are those using a performance-oriented pathways approach, working to quantifiably measure the effectiveness of various strategies in achieving well-defined system performance criteria. For example, for Suzhou in China[5], Singapore[6], Central Cebu in the Philippines[7], Karnataka in India[8], and the Hablehroud River basin in Iran[9], modelling of river catchments has been undertaken to assess flood-reduction interventions and water supply performance under a range of climate change, land use and population growth scenarios. Analyses were done to evaluate combinations of interventions such as bioretention swales, porous paving, green roofing, rain tanks, drain widening, diversion canals, retention ponds, additional groundwater wells and boreholes, new dams, and desalination plants. Combinations of options were assessed against performance criteria under various scenarios to identify robust, flexible, and cost-effective pathways.

Others have taken a less computational, more stakeholder-oriented CRDPs approach primarily aimed at engaging diverse actors in considering and prioritising various development options according to how climate-resilient and carbon intensive they are. For example, in Indonesia's Nusa Tenggara Barat Province[10] and Nuwakot in Nepal[11], researchers have facilitated participatory processes with government and community representatives to collectively determine development objectives and evaluate options against experienced and projected climate risks. In addition to considering physical risks, social drivers of vulnerability were also identified, and options sought that address patterns of economic marginalisation and social exclusion. These options include facilitating access to finance for women-led cooperatives and supporting subsistence farmers to become semi-commercial farmers through increasing market access, improving marketing, introducing new seed varieties, and training on pesticide use.[11]

We chose to focus our review on pathways approaches being implemented in Global South contexts, but there is also a substantial body of literature on efforts implementing adaptation and climate-resilient development pathways in the Global North, notably in the Netherlands, Australia, New Zealand and the United Kingdom. Taken together, this growing body of work offers new insights into concepts of thresholds, signals, decision triggers, lead times between decisions and full operation, path dependencies and transitions from a complex systems perspective, all of which can help guide decision-making under deep uncertainty.[12] One of the challenges identified across many of the studies is how to ensure equity is kept front and centre when applying pathways approaches and working to drive change towards just and inclusive pathways of sustainable and climate-resilient development. The tendency is for dominant actors and the priorities of powerful incumbents to continue shaping the agenda, such that pathways perpetuate marginalisation and disempowerment of those without a large stake in the market economy, such as those practising subsistence farming or engaging in informal urban livelihoods like waste reclamation and recycling.[13] If CRDPs are to stimulate a just transition, then the conception and design of the pathways, and the decisions regarding inevitable trade-offs, need to be inclusive and just in their process and outcomes. This requires negotiating and assimilating (as resolving is unlikely) fundamental disagreements over what constitutes well-being and evaluating effective means of achieving well-being for all.

Shifting South Africa's development pathways

As municipalities struggle to recover from crippling floods, droughts, fires, COVID-19, load shedding, service delivery protests and political instability, and international investors crowd into the alternative energy space in South Africa (including manufacturing green hydrogen and mining manganese and other minerals and metals for batteries), we need to think holistically about development pathways. We need to be designing and investing in development pathways that are attuned to what we carry with us from the past and to the multiple possible futures we are preparing for and bringing into being with the choices we make and actions we take, or shape. How do we combine climate adaptation and mitigation investments into development pathways that give everyone opportunities for a healthier, more sustainable life, especially for those most marginalised to realise their aspirations, including those living with disabilities[14]? A development pathway that is climate-resilient involves a sequence of public and private actions or interventions that steer or nudge the emerging development trajectory towards a safer, more equitable and less carbon intensive future. For example, this might include investing in community-based early warning systems for extreme weather events, planting more drought-tolerant crops, rehabilitating coastal ecosystems to buffer storm surges, building green hydrogen infrastructure, reusing wastewater, employing local teams to clear solid waste and invasive plants from waterways, and so much more. Many of these actions are associated with work opportunities. The key is reconfiguring financial models to make them sustainable. CRDPs are place-based and context-specific with interactions across various spatial and temporal scales, requiring long-term systems thinking for coherent decision-making and coordinated actions.

Many efforts to foster climate-resilient development are already underway in South Africa, as well as elsewhere, but they remain fragmented. We need to continue building bridges between the various pathways-related initiatives to strengthen an integrative, inclusive and evidence-based approach. This requires strengthening the shared evidence base of localised emissions sources, mitigation measures, climate hazards, exposure, sensitivity, adaptive capacity, coping and adaptation measures, and impacts. For example, evidence of changing rainfall intensities and wet spell durations across a municipality, spatial data of locations within the municipality exposed to inundation and erosion, characteristics of households, businesses and infrastructure highly sensitive to inundation or erosion if certain thresholds are exceeded, measures taken to divert floodwaters and protect property, people and animals, and details of flood-related impacts experienced and associated costs. In addition to physical damages, this evidence needs to extend to psycho-social dimensions, such as the psychological burden of financial pressures after widespread crop losses or the social consequences of living in a temporary shelter for months after losing one's home to floodwaters. More aggregated and accessible evidence of the measures taken to reduce climate-related impacts and minimise emissions, with the associated costs, benefits and capacities required to implement and maintain the measures, helps build the case for similar investments elsewhere and weighs up potential synergies and trade-offs between options. For example, there are potential synergies between initiatives that remove solid waste blockages from rivers (to reduce flooding impacts) and those that support local entrepreneurs making products (like outdoor pavers) from reclaimed waste materials. Potential trade-offs need to be suitably appraised against policy objectives, local priorities and system resilience; for example, dam building to reduce water scarcity and drought risk in one area that results in further increasing water scarcity downstream, or building seawalls to protect commercial infrastructure, like hotels, that worsens erosion and inundation further along the coast.

Exciting efforts are already underway working to translate the CRDPs guidance into operational plans through extensive multi-stakeholder processes, for example in the OR Tambo District and Nelson Mandela Bay Municipality in the Eastern Cape, as well as sectoral adaptation pathways being developed for the Western Cape. There is rich learning emerging from these efforts and communities of practice forming between committed change-makers seeking to learn together and support each other, across government, consultancies, academia and civic organisations.

Conclusion

CRDPs are an attempt, bubbling up in various places around the world, to bring together the sustainable development agenda with the climate change and disaster risk reduction agendas, linking bottom-up initiatives with large-scale planning efforts.[15] CRDPs provide a means of coordinating the diversity of actions required to transition to a more just, inclusive and sustainable development trajectory that keeps the global climate system within a range suited to human habitation as we know it.

The review piece published in *Current Opinion in Environmental Sustainability* is an extension of work done with the PCC to develop guidance on co-creating climate-resilient development pathways in South African contexts. The guidance includes a piece on climate information needs and services for developing and navigating pathways, and on the organisational and institutional capacities required to enact a CRDPs approach. The work to build these capacities and co-produce the necessary evidence base is underway, but there is much to be done and learnt. Municipal and provincial governments are on the frontlines of having to allocate resources, deliver public services and respond to extreme events – like floods, heatwaves and droughts – in ways that protect the most vulnerable and enable the private sector to function and grow. Many are struggling and failing to fulfil existing functions, without the burden of increasing climate risks. Coordinated support from the scientific community to understand and monitor climate-related risks and evaluate the efficacy of various adaptation and mitigation measures is needed. Municipalities and the private sector hold some of the data or means of data collection needed to make such analyses possible. The PCC, together with relevant governmental committees and forums formalised through the *Climate Change Act*, provides the institutional architecture through which to cohere such initiatives needed to navigate more climate-resilient development.

The CRDPs approach provides a lens through which to weigh up and sequence development choices and interventions to navigate the choppy waters ahead and can act as a kind of glue to bond together what we are doing with what others are doing, across South Africa and beyond. CRDPs can be mobilised to weave together practical, experiential, indigenous, technical, scientific and policy knowledge to iteratively assess and monitor risks, evaluate and sequence combinations of options, and link locally grounded initiatives with macro-economic interventions, extending across election cycles into multi-generational timescales. Using the CRDPs approach requires tackling power imbalances, mistrust, short-termism, self-centredness, and organisational silos. This is where the role, capacities and convening power of the PCC is key and where academia has a role to play too. We need to move the thinking and science forward in ways that ensure democratic and corporate decisions are being shaped by robust evidence and that actions are being taken that get us closer to realising the goals laid out in so many of our progressive policies, nationally and internationally.

Acknowledgements

I thank the colleagues who were involved in the underlying work referred to in this Commentary, notably Dr Nadine Methner, Kalia Barkai, Prof. Mark New, Dr Christopher Jack, Alice McClure, Penelope Price, Yasirah Madhi, Anna Steynor, Prof. Gina Ziervogel and Prof. Bruce Hewitson, and colleagues in the Secretariat of the Presidential Climate Commission who engaged in the design and development of the CRDPs guidance and case studies. I also thank the anonymous readers for their constructive critiques. Financial support for some of the underlying work was provided by the European Union's Partnership Instrument and the German Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU) in the context of the International Climate Initiative (IKI). The contents of this Commentary are the sole responsibility of the author and do not necessarily reflect the views of the funders.

Declarations

I have no competing interests to declare. I have no AI or LLM use to declare.

References

1. Schipper E, Revi A, Preston B, Carr E, Eriksen S, Fernandez-Carril L, et al. Climate resilient development pathways. In: Pörtner H-O, Roberts DC, Tignor M, Poloczanska ES, Mintenbeck K, Alegría A, et al., editors. Climate change 2022: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press; 2022. p. 3056. https://doi.org/10.1017/9781009325844
2. Werners SE, Sparkes E, Totin E, Abel N, Bhadwal S, Butler JRA, et al. Advancing climate resilient development pathways since the IPCC's fifth assessment report. Environ Sci Policy. 2021;126:168–176. https://doi.org/10.1016/j.envsci.2021.09.017
3. Butler J, Wise R, Meharg S, Peterson N, Bohensky E, Lipsett-Moore G, et al. 'Walking along with development': Climate resilient pathways for political resource curses. Environ Sci Policy. 2022;128:228–241. https://doi.org/10.1016/j.envsci.2021.11.020
4. Taylor A, Methner N, Barkai KR, McClure A, Jack C, New M, et al. Operationalising climate-resilient development pathways in the Global South. Curr Opin Environ Sustain. 2023;64, Art. #101328. https://doi.org/10.1016/j.cosust.2023.101328
5. Xu T, Li K, Engel B, Jia H, Leng L, Sun Z, et al. Optimal adaptation pathway for sustainable low impact development planning under deep uncertainty of climate change: A greedy strategy. J Environ Manag. 2019;248, Art. #109280. https://doi.org/10.1016/j.jenvman.2019.109280
6. Manocha N, Babovic V. Development and valuation of adaptation pathways for stormwater management infrastructure. Environ Sci Policy. 2017;77:86–97. https://doi.org/10.1016/j.envsci.2017.08.001
7. Gilroy K, Jeuken A. Collaborative risk informed decision analysis: A water security case study in the Philippines. Clim Serv. 2018;11:62–71. https://doi.org/10.1016/j.cliser.2018.04.002
8. Bhave AG, Conway D, Dessai S, Stainforth DA. Water resource planning under future climate and socioeconomic uncertainty in the Cauvery River Basin in Karnataka, India. Water Resour Res. 2018;54:708–728. https://doi.org/10.1002/2017WR020970
9. Babaeian F, Delavar M, Morid S, Srinivasan R. Robust climate change adaptation pathways in agricultural water management. Agric Water Manag. 2021;252, Art. #106904. https://doi.org/10.1016/j.agwat.2021.106904
10. Butler J, Suadnya W, Yanuartati Y, Meharg S, Wise R, Sutaryono Y, et al. Priming adaptation pathways through adaptive co-management: Design and evaluation for developing countries. Clim Risk Manag. 2016;12:1–16. https://doi.org/10.1016/j.crm.2016.01.001
11. Pandey A, Prakash A, Werners SE. Matches, mismatches and priorities of pathways from a climate-resilient development perspective in the mountains of Nepal. Environ Sci Policy. 2021;125:135–145. https://doi.org/10.1016/j.envsci.2021.08.013
12. Sparkes E, Werners SE. Monitoring, evaluation and learning requirements for climate-resilient development pathways. Curr Opin Environ Sustain. 2023;64, Art. #101329. https://doi.org/10.1016/j.cosust.2023.101329
13. Singh PK, Chudasama H. Pathways for climate resilient development: Human well-being within a safe and just space in the 21st century. Glob Environ Change. 2021;68, Art. #102277. https://doi.org/10.1016/j.gloenvcha.2021.102277
14. Eriksen S, Grøndahl R, Sæbønes A. On CRDPs and CRPD: Why the rights of people with disabilities are crucial for understanding climate-resilient development pathways. Lancet Planet Health. 2021;5:e929–e930. https://doi.org/10.1016/S2542-5196(21)00233-3
15. Sparkes E, Totin E, Werners SE, Wise RM, Butler JRA, Vincent K. Adaptation pathways to inform policy and practice in the context of development. Environ Sci Policy. 2023;140:279–285. https://doi.org/10.1016/j.envsci.2022.12.011

AUTHORS:
Robert W. Gess[1,2]
Caitlin Reddy[2]

AFFILIATIONS:
[1]Albany Museum, Rhodes University, Makhanda, South Africa
[2]Geology Department, Rhodes University, Makhanda, South Africa

CORRESPONDENCE TO:
Robert Gess

EMAIL:
robg@imaginet.co.za

HOW TO CITE:
Gess RW, Reddy C. South Africa's oldest ophiuroid fossils provide rare insights into the origination of the Malvinoxhosan Realm. S Afr J Sci. 2024;120(11/12), Art. #18896. https://doi.org/10.17159/sajs.2024/18896

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
ophiuroid, Devonian, Malvinoxhosan, Baviaanskloof formation, Gondwana

FUNDING:
South African Millennium Trust, GENUS, South African Department of Science and Innovation, South African National Research Foundation

PUBLISHED:
04 December 2024

South Africa's oldest ophiuroid fossils provide rare insights into the origination of the Malvinoxhosan Realm

Significance:

Recent description of the oldest recorded brittle stars (ophiuroids) from the southern hemisphere revealed two distinct taxa of early Pragian age (approximately 410 million years old). Whereas one of the brittle stars belonged to a species previously described from younger strata, the others represent an unusually spiny taxon formerly unknown to science. Recovered from the 'upper member' of the Baviaanskloof Formation (the uppermost unit of the Table Mountain Group), they offer insights into the earliest recorded phase of the endemic Malvinoxhosan Realm, deposited in polar regions of the ancient supercontinent of Gondwana, and better known from the overlying Bokkeveld Group.

Palaeontology gives us windows into the past and is our only source of direct evidence for the history of life on earth, together with the evolution of the biosphere. It elucidates the increasing complexity of life systems and their response to changing continental configurations, resultant ocean currents, climatic shifts and extinction events. Some of the latter, including the End Devonian Extinction Event, are widely believed to, in turn, result from the evolution of novel biota.[1]

Our view of the past is, however, extremely patchy, often consisting of disconnected glimpses of life in different places at different times. It relies not only on the burial of organisms within suitable sediments, but also on the ultimate burial of those sediments in the earth's crust and their re-exposure at the surface during the brief tenure of current academia.

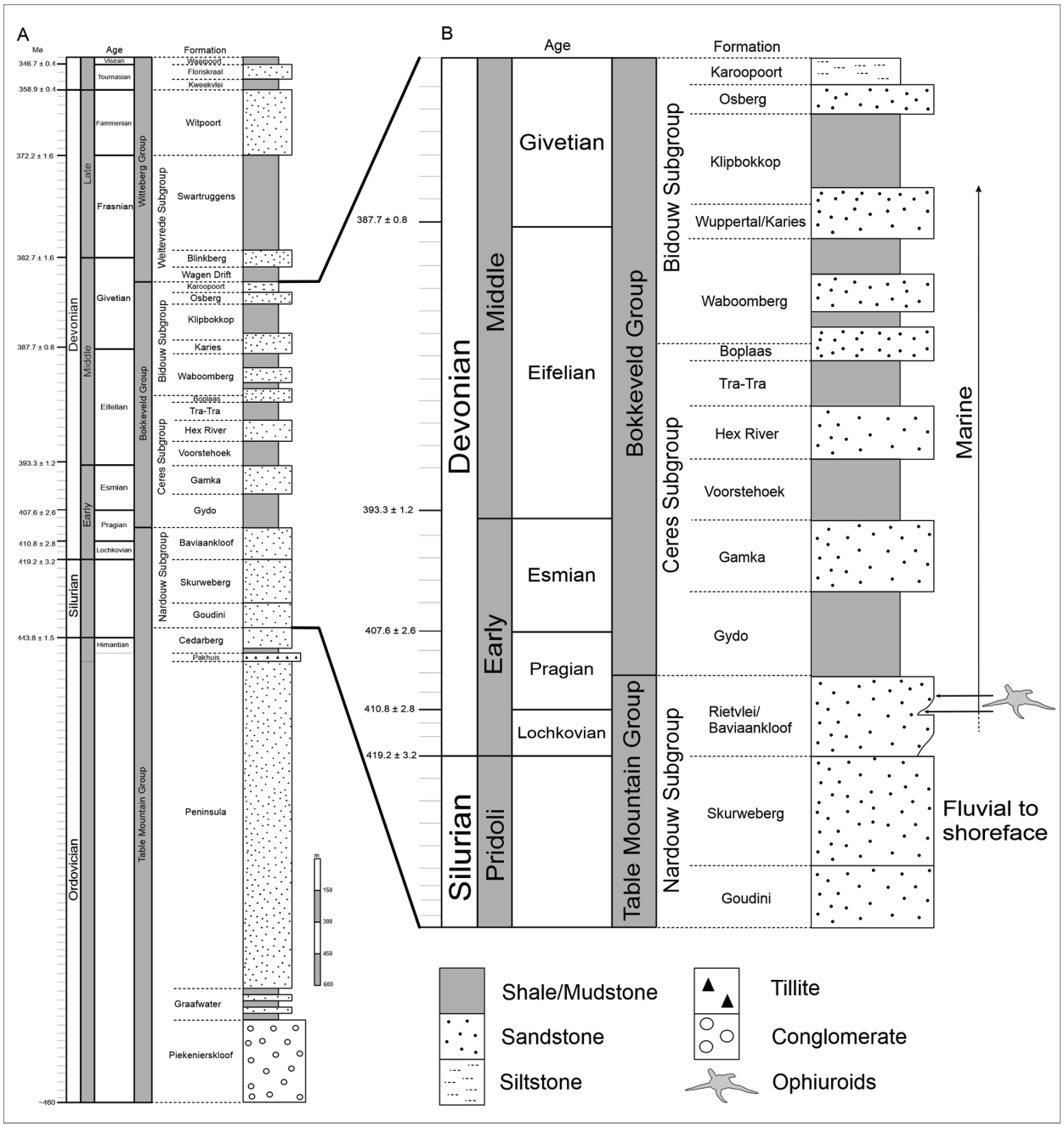
The evidence that we rely on therefore consists of patches of knowability within a matrix of temporal and spacial unknowns. As palaeontologists, we strive to expand the scope of what we know and drive back the boundaries of the unknown.

South Africa's earliest rocks from the Phanerozoic Eon, debatably ranging from the Middle Ordovician Period (around 460 million years ago) until the Early Devonian Period (around 410 million years ago), constitute the Table Mountain Group, the oldest of three major subdivisions comprising the Cape Supergroup (Figure 1).[2] Sediments were deposited in a number of rift generated sedimentary basins, largely by river systems, with occasional marine incursions.[2] Most important of these basins was the Agulhas Basin, which would persist until late in the Carboniferous Period, accumulating a near continuous sedimentary sequence, which also includes the overlying Bokkeveld and Witteberg groups of the Cape Supergroup (Figure 1).[3] These strata, extending parallel to the coast from the Cederberg in the Western Cape to Fish River in the Eastern Cape, preserve an unparalleled record of early Palaeozoic life from the southern polar region of what was then Gondwana (a supercontinent which much later split into Africa, South America, Antarctica, Australia and India). The sandstone dominated Table Mountain Group, represents the Cape Supergroup's initial primarily fluvial stage.[2] As very little continental life existed during these time periods, this interval contains very few fossils. Trace fossils, in the form of invertebrate movement and feeding traces, do occur at various horizons, often associated with episodic marine incursions. The early Late Ordovician Graafwater Formation (beautifully exposed along Chapman's Peak drive in the Western Cape) exhibits a wide range of such ichnotaxa, as do sparse localities within the overlying Peninsula and Pakhuis formations.[2] A single body fossil containing formation within this time interval comprises the Late Ordovician Cedarberg Formation.[2] Cedarberg Formation fossils are largely concentrated within the Soom Shale Lagerstätte, a fine-grained clay horizon deposited during brief flooding of the Agulhas Basin in the immediate aftermath of the Hirnantian glaciation. Characterised by exceptional preservation of their soft tissues in clay minerals, the fossils represent a low-diversity cold-water community including giant conodonts, unarmoured jawless fish, lobopods, brachiopods, orthoconic nautiloids and trilobites.[4] Associated plant spores provide southern Africa's first evidence for terrestrial plants.[5] Low diversity plant spores are also recovered from the more sparsely fossiliferous overlying Disa Member of the Cedarberg Formation, tentatively assigned to an earliest Silurian age on the basis of brachiopod fauna.[6]

A far more substantial fossil record characterises the overlying Bokkeveld and Witteberg Groups (Figure 1), deposited after subsidence and extensive marine flooding of the basin initiated during the early Devonian period. Although first evidenced in the Baviaanskloof Formation (and western equivalent Rietvlei Formation) of the uppermost Table Mountain Group[7,8], this flooding resulted in a fully marine depositional setting for the mudstone dominated Ceres Subgroup (lower Bokkeveld Group), estimated to range in age from the mid-Pragian (approximately 409 million years ago) to late in the Eifelian (approximately 388 million years ago).[9]

The Ceres Subgroup contains an abundant and diverse record of endemic cold-water marine invertebrate faunas, also characteristic of contemporary strata of South America, west Antarctica and the Falkland Islands. These are largely associated with sediments laid down in lower shoreface, offshore transition zone and offshore environments below wave base[9] and form a biotic realm that can be traced upwards as high as the Waboomberg Formation, lowermost formation of the upper Bokkeveld Bidou Subgroup of late Eifelian to earliest Givetian age.[9]

The well-known fauna characteristic of the lower Bokkeveld Ceres Subgroup (and Waboomberg Formation) is traditionally referred to as the Malvinokaffric Realm, in acknowledgement of its commonality with faunas recovered from the Falkland Islands (Malvinas), South America and west Antarctica, although a name change to Malvinoxhosan



Source: Created by Shawn Johnstone, reproduced under a Creative Commons Attribution 4.0 International Licence.

Figure 1: Stratigraphic column of (A) the Cape Supergroup and (B) upper Table Mountain Group and Bokkeveld Group, illustrating the stratigraphic context of the recently described high-latitude Brittle stars.[8,9–11]

bioregion has been introduced in response to sensibilities regarding the word 'Kaffraria'.[12] A huge diversity of marine invertebrate fossils are recorded from this realm, representing 142 genera, including 41 molluscs (of which 24 are bivalves, 8 are gastropods and 3 cephalopods), 34 brachiopods, 36 echinoderms (including 10 crinoids, 2 asteroids/starfish, 6 ophiuroids/brittle stars and 18 of other classes), 15 trilobite arthropods, 6 ostracod arthropods, 9 cnidarians (including 5 anthazoans/corals and 4 conularias), 1 poriferan (sponge), as well as a number of ostracods, hyolithids and tentaculitids.[9] Of these, 37 genera are entirely endemic to the Malvinoxhosan realm.[9] Changing depositional environments within the Agulhas Basin, combined with global changes in climate and oceanic currents, probably resulted in a collapse of the Malvinoxhosan realm during the Givetian, with subsequent faunas belonging to a post-Malvinoxhosan Realm, characterised by rapidly declining representation of endemic genera, an increased proportion of cosmopolitan taxa, and generally reduced diversity.[9]

The origination of the Malvinoxhosan Realm presents a tantalising research question. Fossil bearing marine sediments are largely absent from the South African geological sequence in the interval between strata hosting the distinctive Ordovician cold-water fauna of the Cedarberg Group (Table Mountain Group) and those hosting the early to mid-Devonian Malvinoxhosan faunas of the Early Devonian lower Bokkeveld Group (Figure 1A). Between these lies the Nardouw Subgroup of the Table Mountain Group. Its lowermost subdivision, the Goudini Formation, comprises sandstones of fluvial and nearshore origin, entirely lacking in known body fossils.[13] This is overlain by massive, entirely unfossiliferous fluvial to shoreface quartzites of the Skuweberg Formation which comprises the middle unit of the Nardouw Subgroup.[14] The uppermost unit of the Nardouw Subgroup (and of the Table Mountain Group), formerly known as the 'Passage Beds'[15] is the Baviaanskloof Formation and its western lateral equivalent, the Rietvlei Formation.[3] This Formation records the transition from the fluvially dominated Table Mountain Group

to the marine dominated Bokkeveld Group (Figure 1B). It is informally subdivided into a 'lower member', a middle Kareedouw Member and an 'upper member'. The Kareedouw Member preserves apparent paralic estuarine black mudstones rich in terrestrial plants, although no marine invertebrates. The diverse flora of this unit suggests a Lochkovian (earliest Devonian) age.[8]

However, the Baviaanskloof Formation's upper member contains thin storm deposits of nearshore brachiopods and a low diversity of other taxa. These include the earliest records of Malvinoxhosan indicator taxa, and the combined fauna of the unit has been defined as the 'Eo-Malvinoxhosan'.[9] Prior to the ophiuroids recently described from this unit, only 13 taxa of invertebrates had been recorded, including 1 taxon of bivalve, 1 gastropod, 9 brachiopods, an indeterminate sponge and crinoid remains. All but three of these are also known from the overlying 'Established Malvinokaffric Realm', with the majority being Malvinoxhosan endemics.[9] This suggests that much of the diversity seen in the overlying Bokkeveld Group may have a deeper history, 'off-stage' as it were, in deeper marine settings that are not recorded within our stratigraphic sequence.

In the Eastern Cape, the upper member largely represents shoreface and foreshore deposits[16] containing fairly monotaxic, well-sorted, storm deposits of shallow water *Proboscidina* brachiopods (personal observations of R.W.G.), that, although endemic, are not found in the Bokkeveld Group[9]. Small, highly informative windows into greater diversity are occasionally encountered. A silty to muddy interval, indicative of greater transgression, occurs just above the base of the upper member in the Mpofu Dam section (near Humansdorp). This reveals a bed of shells of *Australocoelia*, a classic and near ubiquitous endemic taxon of the 'Established Malvinoxhosan Realm' – indicating their generally unrecorded proximity in adjacent, slightly deeper water, marine habitats. Approximately 50 cm above this horizon, evidently buried whilst sheltering in a thin storm deposit of *Proboscidina* shells, a single entire ophiuroid was discovered during fieldwork by R.W.G. (Figure 2). Taxonomic work published in Reddy et al.[16] establishes it to be *Hexuraster weitzi* (Spencer, 1950)[17,18], a taxon previously described from the Gydo Formation. This provides the earliest record of an ophiuroid in the southern hemisphere, and again suggests a temporally deeper origination of classic Malvinoxhosan taxa than is normally recorded from our sedimentary record.

Approximately 3.5 kilometres to the southeast of the Mpofu Dam section, at a slightly higher stratigraphic horizon within the upper member of the Baviaanskloof Formation, fieldwork (by R.W.G.) revealed a relatively thick (approximately 10 cm) tempestite lens of *Proboscidina* shells. This deposit, only a few metres across, contained remains of a small population of ophiuroids representing both adults and juveniles. Preserved in life positions, twined amongst the shells (Figure 3), they undoubtedly inhabited the sheltered environment prior to being smothered by a pulse of mud washed out of a nearby river mouth. Bearing thorn-like spines on their dorsal disc and running the length of their arms, they were revealed to be an entirely new genus and species, *Krommaster spinosus* (Reddy et al., 2023)[16], now known only from this single lens of rock.

These windows into the pre-Bokkeveld Malvinoxhosan Realm suggest that the marine faunas typical of the lower Bokkeveld Group were already present during deposition of the uppermost Table Mountain Group, though at slightly deeper water depth than normally recorded in the sedimentary sequence. Furthermore, the discovery of a unique taxon in a single serendipitously preserved lens demonstrates that much of the early, shallow water diversity of the Malvinoxhosan Realm may yet await discovery. The low diversity of marine invertebrates so far described from the Eo-Malvinoxhosan may be ascribed to a number of factors. Firstly, preservation is far rarer than in the fossil-rich shallow marine muds of the lower Bokkeveld Group, including only occasional glimpses of life beyond the low diversity foreshore and shoreface environment. Secondly, exposure of fossil sites is less frequent, the resistant hosting sandstones tending to form weathered hilly country in which softer fossiliferous lenses are destroyed near the surface by weathering. In this regard, it is also worth considering that the fossil-rich mudstones of the lower Bokkeveld Group are a preferred source of soft aggregate for surfacing untarred roads – leading to frequent exposure of fresh material in borrow pits and quarries, from which much of the historically collected material originates. Clearly, more targeted fieldwork

Photo: Robert Gess

Figure 2: Photo of an approximately 10 cm diameter specimen of *Hexuraster weitzi* (Spencer, 1950)[17,18] in situ in strata of the Upper Member of the Baviaanskloof Formation (Table Mountain Group).

Photo: Caitlyn Reddy

Figure 3: Photo of an approximately 10 cm diameter specimen of *Krommaster spinosus spinosus*[16] recovered in life position from a brachiopod lag deposit in the Upper Member of the Baviaanskloof Formation (Table Mountain Group).

within the Baviaanskloof and Rietvlei Formations is required to locate and study the rare palaeoecological windows that they present, in order to better understand the emergence of the Malvinoxhosan Realm.

Acknowledgements

We thank Shawn Johnstone for drafting Figure 1.

Funding

This research forms part of the work of the Devonian Ecosystems Project, which is funded by Millennium Trust, South Africa, GENUS (DSI-NRF Centre of Excellence in Palaeosciences) and the National Research Foundation of South Africa.

Declarations

We have no competing interests. We have no AI or LLM use to declare. Both authors read and approved the final manuscript.

References

1. Algeo TJ, Scheckler E, Maynard JB. Effects of the middle to late Devonian spread of vascular land plants on weathering regimes, marine biotas, and global climate. In: Gensel PG, Edwards D, editors. Plants invade the land: Evolutionary and environmental perspectives. New York: Columbia University Press; 2001. p. 213–236. https://doi.org/10.7312/gens11160-013
2. Penn-Clarke CR, Browning C, Harper DAT. The Ordovician system of South Africa: A review. Geol Soc Spec Publ. 2023;533:175–197. https://doi.org/10.1144/SP533-2022-23
3. Thamm AG, Johnson MR. The Cape supergroup. In: Johnson MR, Anhaeusser CR, Thomas RJ, editors. The geology of South Africa. Johannesburg: Geological Society of South Africa; 2006. p. 443–459.
4. Gabbott SE, Browning C, Theron JN, Whittle RJ. The late Ordovician Soom Shale Lagerstätte: An extraordinary post-glacial fossil and sedimentary record. J Geol Soc. 2016;174:1–9. https://doi.org/10.1144/jgs2016-076
5. Gray J, Theron JN, Boucot AJ. Age of the Cedarberg formation, South Africa and early land plant evolution. Geol Mag. 1986;123:445–454. https://doi.org/10.1017/S0016756800033537
6. Wellman CH, Penn-Clarke CR, Browning C. Early land plant remains from the uppermost Ordovician-?lowermost Silurian Cedarberg Formation of South Africa. Est J Earth Sci. 2023;72(1):164. https://doi.org/10.3176/earth.2023.03
7. Cooper MR. Facies shifts, sea-level changes and event stratigraphy in the Devonian of South Africa. S Afr J Sci. 1986;82(5):255.
8. Gess RW, Prestianni C. An early Devonian flora from the Baviaanskloof Formation (Table Mountain group) of South Africa. Sci Rep. 2021;11(1), Art. #11859. https://doi.org/10.1038/s41598-021-90180-z
9. Penn-Clarke CR, Harper DAT. The rise and fall of the Malvinoxhosan (Malvinokaffric) bioregion in South Africa: Evidence for early-middle Devonian biocrises at the South Pole. Earth-Sci Rev. 2023;246, Art. #104595. https://doi.org/10.1016/j.earscirev.2023.104595
10. Theron JN, Thamm AG. Stratigraphy and sedimentology of the Cape Supergroup in the Western Cape. In: Guidebook, Geocongress '90. Johannesburg: Geological Society of South Africa, PR2; 1990. p. 1–64.
11. Cotter E. Depositional setting and cyclic development of the lower part of the Witteberg Group (Mid- to Upper Devonian), Cape Supergroup, Western Cape, South Africa. S Afr J Geol. 2000;103(1):1–14. https://doi.org/10.2113/103.1.1
12. Penn-Clarke CR, Harper DAT. Early Middle Devonian brachiopod provincialism and bioregionalization at high latitudes: A case study from southwestern Gondwana. Geol Soc Am Bull. 2021;33:819–836. https://doi.org/10.1130/B35670.1
13. Malan JA, Theron JN, Hill RS. Lithostratigraphy of the Goudini formation (Table Mountain group). Lithostratigraphic Series, No. 2. Pretoria: Council for Geoscience, South African Committee for Stratigraphy; 1989. p. 5.
14. Theron JN, Malan JA, HILL RS. Lithostratigraphy of the Skurweberg formation (Table Mountain group). Lithostratigraphic Series, No. 3. Pretoria: Council for Geosciences, South African Committee for Stratigraphy; 1989. p. 6.
15. Rossouw PJ, Meyer EI, Mulder MP, Stocken CG. Die geologie van die Swartberge, die Kangovallei en die omgewing van Prins Albert [The geology of the Swartberg, the Cango Valley and the vicinity of Prince Albert], K.P. Exploration Sheets 3321 B (Gamkapoort) and 3322 A (Prince Albert). Port Elizabeth: Geological Survey of South Africa; 1964. Afrikaans.
16. Reddy C, Thuy B, Reid M, Gess R. Earliest known ophiuroids from high palaeolatitude, southern Gondwana, recovered from the Pragian to earliest Emsian Baviaanskloof formation (Table Mountain group, Cape supergroup) South Africa. PLoS ONE. 2023;18(10), e0292636. https://doi.org/10.1371/journal.pone.0292636
17. Spencer WA. Monograph of the British Palæozoic Asterozoa. Part IX. Palaeontogr Soc Monogr. 1934;87(394):437–494. https://doi.org/10.1080/02693445.1934.12035629
18. Jell P, Theron J. Early Devonian echinoderms from South Africa. Mem-Qld Mus. 1999;43(1):115–200.

AUTHORS:
Nthabiseng Mkhatha[1]
George Djolov[1]
Diego Iturralde[1]

AFFILIATION:
[1]Statistics South Africa, Pretoria, South Africa

CORRESPONDENCE TO:
Diego Iturralde

EMAIL:
Diegol@statssa.gov.za

HOW TO CITE:
Mkhatha N, Djolov G, Iturralde D. Response to Moultrie and Dorrington (2024): 'Problems and concerns with the 2022 South African census'. S Afr J Sci. 2024;120(11/12), Art. #20465. https://doi.org/10.17159/sajs.2024/20465

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
census, South Africa, population, housing, digital census

PUBLISHED:
04 December 2024

Response to Moultrie and Dorrington (2024): 'Problems and concerns with the 2022 South African census'

Statistics South Africa welcomes the opportunity to comment on the Commentary entitled 'Problems and concerns with the 2022 South African census'.[1] The population and housing census (henceforth referred to as 'Census') takes place in South Africa every 10 years and represents a rich source of statistical information that is designed to guide planning and policy development as well as to guide sampling design for the next inter-censal period. Contrary to previous censuses in South Africa, the 2022 Census[2] was South Africa's first digital census. A multi-modal approach was taken by collecting data using Computer Assisted Web Interviewing (CAWI), Computer Assisted Telephone Interviewing (CATI) as well as in person with a digital data collection instrument, Computer Assisted Personal Interviewing (CAPI), and data collection hinged on a geographic digital frame. Whilst all censuses were de facto, including Census 2022, the 2022 Census enumeration period extended over a 4-month period from February to May 2022.

Stats SA, and by extension the Census, was not spared from the impact of the global COVID-19 pandemic. The planning of this Census was severely disrupted by the pandemic, which impacted not only the data collection phase from February 2022 but also the preparatory phases when training and the pilot census were meant to take place. During this preparatory period, the much-publicised KwaZulu-Natal and Gauteng social unrest also took place during the winter of 2021, which restricted the training and pilot related data collection that could take place at that time.

The pilot was eventually executed in only seven provinces. Training reverted from face-to-face to online mode as a result of the COVID-19 restrictions, and, as such, implications for the quality of training were inevitable. Indeed, the training for the pilot had to be suspended in December 2020, when, despite precautionary measures, delegates at a training venue tested positive for COVID-19, and the pilot, due to start in February 2021, was postponed. From a human resource perspective, the uptake of positions for census data collection and supervision was lower than anticipated. As previously indicated, the timelines for CAWI and CAPI were extended due to poor uptake from communities with fieldworkers. Measures to improve uptake were undertaken; these included reallocation of fieldworkers, remote monitoring of data collection, etc.

The digitisation of the census data collection allowed for the enhancement of quality of the processing of Census 2022. Digitisation enabled data validation to be built into the data collection instruments, resulting in minimal data editing and imputations. As part of the census undertaking, a Post-Enumeration Survey (PES) is conducted, which enables not only the quality of the census data to be assessed but also to estimate over or undercount and to adjust accordingly. Through this process, content error and coverage error are assessed. Thus, the PES is a statistically grounded basis for adjusting census counts appropriately. The PES estimated a *final* net undercount of persons of 31.06% and a net household undercount of 30.49%, which was announced at the release of the Census results in October 2023.

Whilst Stats SA forms part of the global population and housing census programme by adhering to the United Nations Principles and Recommendations of Population and Housing Census, a statutory body is in place to ensure that all processes have been followed correctly. The South African Statistics Council plays an important part in the delivery of a census. The South African Statistics Council is appointed by the Minister in the Presidency to exercise oversight over Stats SA and provide counsel to the Minister and the Statistician-General. This oversight involves ensuring that all processes pertaining to the Census have been meticulously followed and requesting updates on progress from the leadership of Stats SA. Facilitating engagement with the data, an independent consultant appointed by the South African Statistics Council critically reviews the implementation of plans and outcomes of operations, rigorously interrogates the data, and ultimately delivers their professional opinion and concerns. This comprehensive process enables the South African Statistics Council to present its informed opinion on the Census. In the specific case in question, the South African Statistics Council expressed its satisfaction with the quality of the data, leading to the launch of the census results in October 2023.

Another issue that should be acknowledged was that the Census was not very positively received by various communities, which is reflected in the contact rates that were monitored during enumeration. COVID-19 impacted training and the pilot census, as indicated previously. Leading into the enumeration phase of the Census, COVID-19 also had an indirect effect through the economic repercussions of all the COVID lockdowns that had previously followed. All of this was very divisive during the lockdown period across the political and societal divide.

Issues with the South African Census of 2022

Balancing equation

It needs to be placed on record that Stats SA has always been transparent in methods and operations translating to generated official statistics. While the article[1] cites two census evaluation methods, the authors seem to suggest that the balancing equation method they used provides better population estimates when compared to Census 2022. Such an assertion is misleading. Stats SA has equally and consistently used the Post-Enumeration Survey (PES)[3] as part of the census data evaluation. PES is used to determine the degree of coverage errors and improve

the quality of census results through the application of adjustment factors generated from the PES. The PES, undertaken after the conclusion of the Census enumeration, also allows for examination of characteristics of persons that may have been missed during enumeration, as per the guidelines in the UN Principles and Recommendations (Rev3, par 3, p.221[4]).

The authors of the Commentary must be aware that assumptions on the three demographic processes (fertility, migration and mortality) underlying the balancing equation method vary based on available information at the time of estimation. Such assumptions are the basis for variations in population estimates by different researchers. It is surprising for the authors to seem to suggest that modelled data based on theoretical assumptions is more accurate compared to a census count, adjusted for under coverage. The authors seem to suggest that either their modelled population estimates and Stats SA's mid-year population estimates (MYPE) for the year 2022 are the closest reliable estimates compared to the Census 2022 population count. Important to note is that the MYPE population has its own assumptions. Traditionally, modelled data are updated when a new census data point becomes available. Although there has been upward and higher than usual undercoverage, as highlighted in the PES report in Census 2022, the results are based on the actual count, and transcend the MYPE and any other population estimates. Furthermore, the authors' assessment overlooked the residual (e), which denotes the error of closure. For census evaluation, it is essential to include the residual (e) to balance the equation precisely. In this context, "e" signifies the "error of closure", encompassing the balance of errors in the data on births, deaths, net migration, and the coverage of the two censuses.

Regarding the analysis on the balancing equation, the error in Limpopo is noted from Table 1; however, a further correction of the value for item 6 is that the value for immigration by place of residence should be 1 145 835, which makes the reported/implied migration 30.5%. It is impossible to comment on the reported immigration by place of birth as the assumptions and computation of estimated survivors of foreign born individuals in 2011 are not shared.

It is apparent that the authors undermine Stats SA's technical capabilities in producing reliable, accurate statistics, be it MYPE or censuses. Statements in the article such as "Stats SA methodology for producing their population projections is somewhat opaque" illustrate inherent undermining behaviour that is uncalled-for in an academic and research space.

Comparison with previous results

When comparing the current census to previous ones, it must also be noted that previous censuses have their own considerations and biases, which might distort some comparisons. It is also reasonable to assume that, particularly between 1996 and 2011, the quality of the censuses improved, which makes a comparison between 1996 or 2001 with 2022 biased by different quality measures. The use of population estimates, whether from Stats SA or the SAMRC-UCT collaboration, is noted as an independent data source against which comparisons are made. It must be noted that the different assumptions between the SAMRC-UCT and Stats SA models, and by extension the model derived by any other independent researcher, are likely to be different. One must nonetheless acknowledge that such estimates prior to the 2022 Census do not take into account any changing patterns that the Census might reveal, particularly at lower levels.

Issues at sub-provincial level

Finally, with regard to the issues at the sub-provincial level, although the evidence is convincing, care should be taken to compare with IEC data, which only capture those over 18 and those interested in voting at the local government level. It is also common, albeit anecdotally, that people who have recently moved to a province are often likely to not re-register with the IEC at their newer place of residence and are happy to vote only in general elections. The differences in growth between the IEC data and the Census are nonetheless noted. It must also be noted that satellite imagery should be used with caution, because in built-up areas it is expected that the population may grow in areas where there has been little to no physical growth.

In parting, Stats SA has recently published the 2024 mid-year population estimates by implementing the cohort-component method and using empirical data sources to reconstruct the population in the inter-censal period 2011–2022. The MYPE provides a series of population estimates that are not only robust but also align themselves to the 2001, 2011 and 2022 Censuses. It must further be stressed that we should not lose sight that 70% of the population was counted and that these data are very useful and valuable.

Post-Enumeration Survey

This section seeks to address the following two points:

1. Preliminary undercount rate vs final undercount rate
2. Computation of standard errors for the adjusted population estimates

1. Preliminary undercount rate vs final undercount rate

While Stats SA encourages the public at large and academia, including renowned researchers, to use census data in their areas of interest, at the same time, it cautions all stakeholders to use data responsibly. Using data responsibly includes highlighting data limitations on Census 2022 and any other products the organisation produces that can inform future improvements and enhancement of methods, processes and products. The introductory remarks of the authors quoting preliminary undercount figures for white and Indian/Asian populations instead of actual figures which were readily available points to the misuse of statistics and an attempt to undermine the credibility of Stats SA.

The article does not make a distinction between the preliminary and final undercount rates. The preliminary undercount rate (29.6%) is derived directly from the PES estimate of the population, while the final undercount rate (31.1%) is calculated based on the adjusted and content-edited census data. Therefore, any mention of the official undercount rate should make reference to the final undercount rate.

The article further references undercount rates of 72% and 62% of the Indian/Asian and white population groups, respectively. It should be stressed that these are preliminary undercount rates and should not be mistaken for the final undercount rates. The final undercount rates for these population groups were estimated at 42.10% for the Indian/Asian and 24.86% for the white population groups (see Table 15 in the PES report[3]).

2. Computation of standard errors for the adjusted population estimate

The article makes an assertion that there could be a computational error in the derivation of the standard errors (SE) for the adjusted population estimate in the PES 2022. The basis for this assertion seems to be the expectation that a final undercount rate of 31.1% should inherently result in a higher SE for the adjusted population estimate in Census 2022 compared to Census 2011, given the lower undercount rate of 14.6% in Census 2011.

As it was the case in Census 2011, the adjusted population estimate in Census 2022 is a product of the actual census counts and adjustment factors within homogeneous classes obtained by post-stratification, as detailed in the released PES report. Table 1 shows estimates of the undercount rate and their SE, and estimates of the population and their SE at national and provincial levels for 2011 and 2022. The calculation of SE for the population estimate in PES 2022 has been verified and confirmed as reported. The SE is a measure of how much the estimate would vary if the survey could be repeated over and over using a different sample of units of the same size every time, i.e. it gives an absolute amount of variability in the estimate. The SE for population estimate is dependent on the variation associated with the estimate and not on the magnitude of the undercount. A higher undercount rate does not necessarily translate to a higher SE of the population estimate. In theory, a smaller sample size would result in high standard errors. However, the undercount rates are not sample sizes but estimates of the missed population.

The coefficients of variation (CVs) of the population estimates for 2011 and 2022 are within the prescribed limits of the South African Statistical Quality Assessment Framework. The CVs measure the variability of each estimate

Table 1: Estimates of the undercount rate (UC) and their standard errors (UC SE), and estimates of the population and their standard errors (SE), for South Africa and its provinces, 2011 and 2022

	2011		2022		2011			2022		
	UC (%)	UC SE (%)	UC (%)	UC SE (%)	Population	SE	CV (%)	Population	SE	CV (%)
Western Cape	18.5	0.542	35.56	1.56	5 822 734	40 830	0.70	7 430 000	70 000	0.98
Eastern Cape	12.9	0.196	30.56	0.59	6 562 053	125 810	1.92	7 230 000	60 000	0.86
Northern Cape	13.4	0.318	29.11	2.27	1 145 861	82 466	7.20	1 350 000	30 000	2.20
Free State	10.1	0.362	18.52	2.62	2 745 590	117 567	4.28	2 960 000	40 000	1.35
KwaZulu-Natal	16.7	0.379	31.19	1.08	10 267 300	109 994	1.07	12 390 000	100 000	0.77
North West	14.9	0.532	17.92	5.39	3 509 953	166 754	4.75	3 800 000	40 000	1.14
Gauteng	14.7	0.174	30.79	2.07	12 272 263	106 023	0.86	15 120 000	90 000	0.61
Mpumalanga	15.5	0.473	34.00	4.51	4 039 939	219 299	5.43	5 160 000	60 000	1.13
Limpopo	10.0	0.135	23.77	1.02	5 404 868	251 244	4.65	6 570 000	60 000	0.91
South Africa	**14.6**	**0.132**	**29.6**	**0.82**	**51 770 560**	**997 560**	**1.93**	**62 030 000**	**120 000**	**0.19**

Note: The standard error (SE) is a measure of how much the estimate would vary if the survey could be repeated over and over using a different sample of units of the same size every time, i.e. it gives an absolute amount of variability in the estimate.

The coefficient of variation (CV) or relative standard error (RSE) is the ratio of the standard error of the survey estimate to the value of the survey estimate expressed as a percentage. It gives the relative amount of variability instead of the absolute amount of variability in an estimate. It allows for comparing domains whose estimates and standard errors differ in magnitude.

and, therefore, the CVs for coverage error rates and population estimates cannot be compared with each other as they refer to different constructs.

Finally, the language used in the article, including in its title 'Problems and concerns with the 2022 South African census', as well as the introduction, conclusion sections, and the various subtitles referred to as "issues" in the journal article, is characterised by a sensational and malicious tone. It appears that the article's intent was not solely to highlight data limitations, as is common practice, but to cast doubt on the credibility of Stats SA, a renowned national statistical office. The use of such dramatising language in a journal article is concerning, as it may indicate an underlying agenda to discredit the Census 2022 results. The article suggests that the census was conducted without sufficient technical expertise, thereby calling into question the validity of its operations, methods applications, and data evaluation procedures. This insinuation not only undermines the integrity of Stats SA but also raises doubts about the accuracy of the Census 2022 results.

References

1. Moultrie TA, Dorrington RE. Problems and concerns with the 2022 South African census. S Afr J Sci. 2024;120(7/8), Art. #18585. https://dx.doi.org/10.17159/sajs.2024/18585
2. Statistics South Africa. Census 2022 statistical release. Pretoria: Statistics South Africa; 2023. Report P0301.4. Available from: https://census.statssa.gov.za/assets/documents/2022/P03014_Census_2022_Statistical_Release.pdf
3. Statistics South Africa. Post-enumeration survey statistical release. Pretoria: Statistics South Africa; 2023. Report P0301.5. Available from: https://census.statssa.gov.za/assets/documents/2022/P030152022.pdf
4. United Nations. Handbook on the management of population and housing censuses, revision 2. New York: United Nations; 2021. Available from: https://unstats.un.org/unsd/publication/seriesf/series_f83rev2en.pdf

Note from the Editor-in-Chief:

We are grateful to Stats SA for their response. As with all such items for our Journal, we provided the authors with comments from expert readers; the authors did not change their submission subsequent to receiving these comments, and the response has been published as submitted, in the interest of open debate in science and its real-world applications.

AUTHORS:
Tom A. Moultrie[1]
Rob Dorrington[1]

AFFILIATION:
[1]Centre for Actuarial Research, Faculty of Commerce, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Tom Moultrie

EMAIL:
tom.moultrie@uct.ac.za

HOW TO CITE:
Moultrie TA, Dorrington R. Rejoinder to Mkhatha et al. (2024): Problems and concerns with the 2022 South African census. S Afr J Sci. 2024;120(11/12), Art. #20466. https://doi.org/10.17159/sajs.2024/20466

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
census, demography, South Africa, migration, age-sex distribution

PUBLISHED:
04 December 2024

Rejoinder to Mkhatha et al. (2024): Problems and concerns with the 2022 South African census

We welcome this opportunity to engage Statistics South Africa (Stats SA), even if only on a subset of the substantive issues we have raised[1,2] about the data from the 2022 South African census. Many of the issues were first ventilated in a memorandum sent to the South African Statistician-General in November 2023. Nevertheless, the nature of Stats SA's response still is overwhelmingly one of simply asserting that our concerns are misplaced or unfounded.

The litany of problems mentioned in the first six paragraphs of their response is noted. Each of these could have materially compromised the census results in both predictable and unpredictable ways. Faced with these challenges, it is perhaps not surprising that the undercount in the census was so great. We reiterate that, given the impact of the pandemic and its associated lockdowns on the functioning of all organisations, the decision to implement the census when they did was ill-advised. Even not knowing at the time when the pandemic would have run its course, it was unwise not to delay the census to October 2023, or even October 2024.

The balance equation

Stats SA's response[3] either misunderstands or misrepresents our analysis based on the 'balance equation'. Fundamentally, and wrongly, they suggest that we use the balance equation to determine an alternative set of estimates of the population.

We do not. The analysis permits the identification of inconsistencies arising from the components of the equation (unpacking the "error of closure" to which Stats SA refer). We argue that the magnitude of differences cannot be explained by errors in the numbers of births and deaths, which we clearly acknowledge and point out is, relatively, small, so must be due to either errors in the 2011 census (and by implication also, substantially, the two previous post-apartheid censuses to the extent they are consistent with the data from 2011), and/or migration, and/or the 2022 census results. We note that as the difference/residual between the 2022 census and projections must imply a pattern of migration that is inconsistent with what would be expected (particularly at older ages), the implication is that the 2022 census estimate is too high. (In passing, try as we might and notwithstanding the fact that the published numbers have changed slightly between different published versions of the data, we cannot find any errors in our derivation of the numbers of immigrants from the place of previous residence, and are forced to assume that Stats SA's observation that the number of immigrants derived from these data is over 1 million is an error on their part). However, although we do not suggest that either 'our' or Stats SA's 2022 mid-year population estimate (MYPE) projections are acceptable alternative estimates of the population at the time of the census (both, to different degrees, appear to miss what could be migration of people aged 20–39, for example), we do argue that these projections of, for example, children, particularly aged 5–9, and those aged over 50 make more logical sense.

From this, and without producing an alternative estimate, we do suggest that the 2022 census could have overestimated the true number by "approximately one million". In this regard, it is instructive to note that interpolation from the 2024 release of Stats SA's own MYPEs produces an estimate that is nearly one million (960 000) *below* the census estimate at the census date of 2 February 2022. We note further that there has been a significant change in the MYPEs at provincial and particularly sub-provincial levels since the release of the census results. This has produced a number of curiosities in district-level population estimates. (As an important example, the implied population of the City of Johannesburg in February 2022 is some 16% higher in the new series of MYPEs than revealed in the 2022 census).

Such anomalies feed into the equitable share formulae used to allocate revenue and resources.

We do not presume that a cohort-component population projection from an earlier census will produce a superior estimate of the population at the time of the census than a successfully run census and Post-Enumeration Survey (PES).

Unfortunately, as we have expressed, we have doubts about how successful the 2022 PES was.

The Post-Enumeration Survey and small-population estimates

As we highlighted in the Technical Report[1], it is probable that the sample size of the PES was too small to accurately adjust the actually enumerated population in the census (of 36 million) to the final estimate of 62 million. This manifests in the severe anomalies in the estimates of the population of small groups and in small areas, some of which Stats SA concede is "convincing".

And yet, even in regard to the anomalous estimates we highlight, Stats SA erects strawmen to cast doubt on the issues raised. Of course, the IEC Voters Registration data are imperfect (although we did restrict our analysis to comparisons to the census population over 18) and may not fully capture the extent of migration. However, the population aged 18 and over in the Richtersveld local municipality more than doubled (apparently) between the censuses, while the IEC data from two local government elections held very close to the census dates show an increase of less than 5%. It is difficult for Stats SA to explain such results with generic expressions of (already acknowledged) concerns as to the weaknesses of such comparisons.

These census numbers defy common sense.

Likewise, they strawman the analysis of the Beaufort West town and local municipality, suggesting that satellite imagery of dwellings might hide increased numbers of people living in each dwelling. Hypothetically, that might be so. However, Stats SA's own data (in a report that no longer appears on their website) for the municipality indicate that the average number of people in a household in the municipality was unchanged between 2011 and 2022, at 3.8 persons.[4] In other words, that simply cannot be the explanation because, if the average household size has not changed, and there are insufficient new dwelling units visible in satellite imagery, the purported growth in the municipal population is logically incoherent.

In preparing this response, too, we note that Stats SA is not above practising legerdemain. Their response makes much of the difference between the preliminary and final estimates of the undercount, and goes so far as to insinuate bad faith on our part. However, on the day the census results were released (10 October 2023), we each downloaded all documents and files made public on Stats SA's website, including the report on the PES. In order to respond to Stats SA's rebuttal, we have found that the version of the PES report that we downloaded has been replaced with a newer version, 4 pages longer, and including the tables they mention. At no point did Stats SA alert users to the new version or make it clear that the first version had been superseded.

Having now been alerted to the new tabulations, we find that they amplify our initial concerns. While the national estimate of the undercount shifts from 29.6% to 31.1% between the preliminary and final estimates, that of every population group narrows by between 5 and 37 percentage points. These changes would appear to be due largely to imputation of a very high proportion of people with undetermined/unknown population group enumerated in the census. In this regard, Stats SA has never revealed the extent of imputation of key variables or the processes and heuristics used in imputation, nor have they made data available to users that will allow investigation of the impact of the imputation procedures (by, for example, including edit flags in the 10% sample released to the public in August 2024, as was done for the 2001 census).

Finally, inter alia, probably as a consequence of this imputation, with the release of the 10% sample, we have identified a large number of worrying aspects in the weighting and the construction of that sample. As was the case previously, we will share these with Stats SA before publishing. Extreme deviation of weights in the 10% sample within the clusters used to derive the adjustment factors from the PES reinforces our contention in November 2023 that the data are not fit for purpose.

References

1. Moultrie TA, Dorrington R. The 2022 South African census. Cape Town: South African Medical Research Council; 2024. Available from: https://www.samrc.ac.za/research-reports/2022-south-african-census
2. Moultrie TA, Dorrington RE. Problems and concerns with the 2022 South African census. S Afr J Sci. 2024;120(7/8), Art. #18585. https://doi.org/10.17159/sajs.2024/18585
3. Mkhatha N, Djolov G, Iturralde D. Response to Moultrie and Dorrington (2024): 'Problems and concerns with the 2022 South African census'. S Afr J Sci. 2024;120(11/12), Art. #20465. https://doi.org/10.17159/sajs.2024/20465
4. Statistics South Africa (Stats SA). Census 2022 municipal fact sheet. Report 03-01-82. Pretoria: Stats SA; 2023. Available from: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/982/download/12776

AUTHORS:
Mehita Iqani[1]
Anna Feigenbaum[2]
Laila Asmal[3]
Linsey McGoey[4]
A.M .Kanngieser[5]
Alexander Dunlap[6]
Thema Monroe-White[7]

AFFILIATIONS:
[1]South African Research Chair in Science Communication, Stellenbosch University, Stellenbosch, South Africa
[2]Department of Communication and Journalism, Bournemouth University, Bournemouth, United Kingdom
[3]Department of Psychiatry, Faculty of Medicine, Stellenbosch University, Cape Town, South Africa
[4]Department of Sociology, University of Essex, Essex, United Kingdom
[5]Department of Geography at Royal Holloway, University of London, Egham, United Kingdom
[6]Institute for Global Sustainability, Boston University, Boston, Massachusetts, USA
[7]Schar School of Policy and Government, George Mason University, Fairfax, Virginia, USA

CORRESPONDENCE TO:
Mehita Iqani

EMAIL:
mehita@sun.ac.za

HOW TO CITE:
Iqani M, Feigenbaum A, Asmal L, McGoey L, Kanngieser AM, Dunlap A, et al. Speaking evidence to power? An interdisciplinary conversation. S Afr J Sci. 2024;120(11/12), Art. #19128. https://doi.org/10.17159/sajs.2024/19128

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
evidence, power, interdisciplinary, knowledge production, resistance

Speaking evidence to power? An interdisciplinary conversation

Significance:

Interdisciplinary conversations about how knowledge is produced are significant in that they allow for reflection and exchange between the 'silos' of academic institutions. This Structured Conversation explores some of the complexities that are attendant to generating and analysing data, and how data are then entered into relationships of authority and resistance.

Introduction

Mehita Iqani (M.I.) and Anna Feigenbaum (A.F.): This Structured Conversation presents a synthesis of the first Evidence & Power roundtable webinar, which took place on 15 June 2023. Evidence & Power is a roundtable series co-hosted by the South African Research Chair in Science Communication (Stellenbosch University, South Africa) and the Centre for Science, Health and Data Communication Research (Bournemouth University, UK). The roundtable summarised in this essay was co-convened and co-chaired by the first two authors (M.I. and A.F.), and hosted five scholars (the remaining co-authors) from neuropsychiatry, sociology, geography, anthropology, and data analytics, respectively. What follows is an edited version of the rich discussion that took place in the webinar. The full webinar recording is available on YouTube.

The purpose of this webinar series is to create an iterative, ongoing meta-theoretical discussion about what forms of evidence are created and deployed in various disciplines, how these intersect with power, and how, together, questions of evidence and power shape the academic project in general. These issues are, arguably, central to the politics of what it means to create scientific knowledge (and communicate it) in South Africa and, more broadly, in Africa. Precisely because of the imbalances in the knowledge economy globally, African researchers are often perceived by the North to be on the receiving end of epistemology as well as funding, even though, more often than not, they are at the heart of innovation and new discoveries. This roundtable series inverts that assumption and brings South African and African knowledge-makers to the centre of debates that should be global and inclusive.

The convenors conceptualised this seminar series, and decide who to invite to speak. Our politics is aimed at challenging North-South divides in knowledge production, radical hospitality in terms of cross-disciplinary conversations, and an attention to always ensuring that 'emerging' scholars have opportunities to converse with those more established on an equal platform. This results in a conversation that is wide-ranging and which includes a diversity of depth and breadth of commentary. Specifically, we aim to connect researchers from South Africa and Africa and the rest of the world, who are working in innovative ways at the interface of science, communication, and society to explore and theorise the relations between data, methodology and the exercise of scientific authority. The intentionally cross-disciplinary, trans-career level profile of each roundtable allows for a textured interdisciplinary discussion on the overarching theme.

What counts as evidence?

In the first round of discussion, speakers were asked to reflect on the specific forms that data take in their disciplinary traditions, and to articulate how those data forms are assessed to be valuable to a particular intellectual project.

Laila Asmal (L.A.): In psychiatry, our core focus lies in understanding, diagnosing, treating, and preventing mental disorders. When it comes to data generation, the last century has seen psychiatry taking a fundamentally scientific approach. This involves systematic methodologies in the generation of knowledge, starting with hypothesis formation and progressing through the collection of evidence to test this hypothesis. A crucial step in this process is identifying and addressing potential biases, followed by the iterative process of retesting and refining our hypotheses. This often involves studying diverse sample populations to enhance the robustness and applicability of our findings. The fundamental premise of science is that evidence is replicable and is unbiased (or if there are biases, these biases are acknowledged and are addressed). Science is sometimes described as a competition of ideas, not of people. The evidence it generates needs to be trusted by society. In Western medicine, which is the dominant frame internationally, it is crucial that research is viewed by society as a sincere, unbiased attempt by scientists to describe the world accurately. In psychiatry, we continually navigate the tension among social science perspectives, therapeutic approaches and biological methodologies that often results in a push and pull in both the generation and interpretation of evidence.

Linsey McGoey (L.M.): In qualitative sociology and social theory, one of the objects of study is how evidence is produced and how different political, social, and economic actors use evidence to meet different ends. Questions of power and domination are bound up in the study of evidentiary regimes in practice. In some branches of sociology, quantitative evidence is at the top of the knowledge hierarchy: big data sets are seen as more convincing than qualitative evidence. Ignorance studies consider how the absence of evidence can be a political goal for actors who do not want to be accountable for the negative effect that their actions can have. This is in contrast to a model of power in which the more a state knew about subjects, the more it could control them and potentially dominate them, as work by scholars such as James Scott[1] have emphasised. Ignorance studies show how *not knowing* is also useful as an instrument of power, and how the absence of a fact is evidence itself. In early attempts to publish these findings, there was some resistance to this form of evidence from journal editors and reviewers.

A.M. Kanngieser (A.M.K.): In the sciences, climate change evidence is often drawn through quantitative data sets collated from different sensing technologies over diverse environmental registers. There is also ethnographic work in disciplines adjacent to and within geography on the ongoing impacts of colonialism and ecological disasters in frontline communities in the Pacific.[2,3] Ethnographic accounts tell stories that are different from quantitative data collection that seeks to export case studies for broad application. It is interesting to consider research in this domain that places into dialogue quantitative and qualitative sources to see how these sit together, because they can reveal many tensions. Ethnographic evidence can show how climate change is not a one-size-fits-all phenomenon, and that there are differential impacts happening in different or even the same places. People's experiences of those impacts can contradict what quantitative data project.[4,5] So the question of compelling evidence becomes messy when there are contradictions between long-range global data sets and ethnographic accounts of lived experience. The challenge will be to bring these two forms of evidence together, because the complexity of lived experience can undermine or complicate sensor readings.

Alexander Dunlap (A.D.): In political ecology, it is probably most accepted to start with an understanding that objectivity is impossible; that all epistemology is constructed by discourse[6,7], and all evidence is always going to have some kind of inherent bias or subjectivity. It is concerning that any data, even quantitative models, are passed off as objective.[8] Ecology makes claims to objectivity; political ecology says that it is impossible because politics and ecology are irreparably intertwined.[9,10] What is compelling evidence when we live in a political system that is more or less completely unresponsive to ecological catastrophe, social crisis, and the climate crisis? The notion of evidence being 'compelling' does not hold, because evidence often does not lead to decision-making; economic interest and political power do — whoever has the most money, will initiate extractive projects, pay things off, and keep the existing system going in the way that is suitable to capitalist political economy. Although it may not have the same power to sway, in political ecology, the most compelling evidence comes from being present on the ground, spending time, interviewing, and triangulating all of those data with multiple sources of corroboration.

Thema Monroe-White (T.M-W.): In the field of science, technology, and innovation (STI) policy, quantitative data generated from large quantitative data sets has had an outsized influence on policy audiences. More recently, this quantitative framing has been rebranded under the umbrella term, 'analytics', which relies on applying computational and statistical approaches to data sets for the purpose of generating informed decision-making. Data analytics is not unlike its disciplinary cousins (data science, machine learning, big data, artificial intelligence, etc.) in that it serves as a highly influential tool because it reinforces the long-standing values of quantification that have remained central to principally economic policymaking. In turn, the field of analytics as a whole has been equated with 'compelling evidence'. In STI policy, what counts as compelling evidence is broadly understood to include rigorous and empirical insights that are reproducible and/or transferable to different contexts. Unsurprisingly, these same quantitative studies are also sought out as validation of the empirical evidence generated by qualitative studies and scholars. The aims of reproducibility via open access code and data sets, while highly respectable and desirable, are all too often overshadowed by this prioritisation of the quantifiable (i.e. statistical and computational). In fact, insights derived from this compelling evidence can cause more undue harm if the data-driven analyses fail to be adequately humanised (i.e. critically examined in the context of the lives they reflect and/or are meant to serve). The limits of quantification have yet to be fully recognised by STI policy audiences; that said, the rise of critical quantitative scholarship focused on algorithmic bias, data harms, and methodological pluralism is helping to balance the scales.

How can evidence be used as an instrument of power?

In the second round of discussion, speakers were asked to reflect on how the forms of evidence that tend to dominate in particular disciplinary domains are, or can be, deployed from the top down, as tools of authority, control or oppression.

L.A.: Psychiatry, as a medical specialty, has been dominated by a biological and quantitative approach since the second half of the 20th century. For researchers in low- to middle-income countries, this raises concerns. As we entered the millennium, 90% of the global population lived in low- to middle-income countries, but only 5% of mental health research in the high-impact journals (the ones that actually matter in terms of change in practice) came from those countries.[11] This speaks to the bias towards quantitative research, which goes largely unrecognised. At a very basic level, quantitative research in the medical sciences is expensive. Winning big grants to do such research requires a fair amount of technical know-how, and as such is gamified towards high-income countries. Researchers in low-income countries usually end up with small sample sizes or qualitative data that are not usually regarded as good medical science. Such studies do not hit the level of large quantitative studies. This means that, in Africa, for example, it is not uncommon for African collaborators to be included in big projects led by researchers in high-income countries, and to be simply treated as data collectors. The high-income country researchers prioritise what research questions are asked and the approach, and African collaborators are sidelined. It is not infrequent for a high-income country's collaborator to 'swoop in from the fancy plane' to make broad recommendations, and then the low-income country's researcher needs to do the work on the ground for little credit. In a system in which only 5% of the high-impact journal articles are published with data from low-income countries, researchers from those countries have to toe the line. It's high stakes to talk about systemic inequality in the research funding environment, because the journal editors, the conference organisers, and the people who have access to a lot of the journals (because these journals are expensive) all exercise power within the system.

As an example, an interesting study was done in which the researchers looked at tuberculosis treatment amongst a well-performed randomised control trial, and they had clinicians from England review the study. They did not reveal or explain from where the scientists came, and the clinicians rated the study very highly. When they randomly assigned the publication source to either a high-income or a low-income country authorship, the clinicians rated the article differently.[12] The findings provided evidence that when a study is seen to be from a high-income country, clinicians rate it as more relevant and recommend the finding to peers; if the study was perceived to come from a low-income country – featuring the *exact* same science around tuberculosis treatment – the recommendations were completely different with lower ratings of relevance and lower peer recommendations. This has an impact on the scientific trajectory of scientists: those from low-income countries are likely to have less success with grant applications, to be invited to fewer conferences and so on. It also has implications for knowledge translation and accessibility, in terms of which research findings make it into clinical practice and patient care. This is an accepted and unspoken status quo; however, there are examples of change starting to happen. Low- and middle-income countries provide important and unique insights into long-term outcomes and intergenerational effects of trauma (including post-traumatic stress disorder), HIV and mental health, mental health services research, neurogenetics and neurodevelopment.

L.M.: Another worthwhile case study for thinking about how evidence is used as an instrument of power is a long-term study of the Bill and Melinda Gates Foundation, which I carried out before the wave of recent conspiracy theories related to Gates, COVID-19, and vaccines. The study showed how evidence can be used in unintended ways and how those generating that evidence need to take accountability for that. When the Gates Foundation got involved in global health funding, they directed the vast majority of their research grants towards recipients in developed, wealthy nations.[13,14] Recipients tended to be based at Ivy League institutions, and there was a debate about the justice of this. Speaking up about this carried the risk of losing potential funding from that Foundation, however. When pressed, the Foundation decision-makers would say that science should be a competition of ideas, and not people, and that they chose to fund the most cutting-edge science. Although they claimed to want to end world inequality, surely, they should have tried to enhance the research capacity in developing nations rather than compound existent inequalities. These arguments have largely fallen on deaf ears. Although critiques of the Gates Foundation were valid, unintentionally they also

fuelled extremist perspectives on philanthropic activities in general and bolstered a perception that there are shadowy pockets of people trying to control world events. When a valid critique gets redeployed as a kind of misinformation, it can be very dangerous. Ignorance studies can feed into conspiratorial logic, because it suggests that some of the most important types of power are not visible to certain actors. This in turn can lead to extremist reactions which can fuel violence and inequality. This unintended deployment of evidence as a tool of division is very concerning. It shows that evidence is not something that researchers can ever fully control, nor seek to.

A.K.: In climate research, quantitative data sets are presented as objective – this is impossible. More attention must be paid to how and by whom the data are produced – how data are generalised, despite local differences, from where they are collected, and, very importantly, the geopolitical relationships surrounding the production and use of the data. For many reasons, including ethics and onto-epistemological differences, researchers in the social sciences incorporate and centre ethnographic data where possible, as some kind of counterpoint to quantitative data. An issue here is that, in the Pacific climate space, a lot of the ethnographic work is conducted by white Australian, New Zealand, European, and American researchers. Equally here, there needs to be open self-reflection about who is forming research agendas and themes, who asks questions, the language of delivery, and how research participants are recruited. Furthermore, there is limited transparency about how researchers are funded, their ethics, and with which organisations they do or do not collaborate. Funding is largely allocated to Western researchers coming from outside, and funding grants are written in alignment with the funding government's political and economic principles and interests, rather than those of local Pacific communities. In Australia, for example, there is a strong push for public-private partnerships and industry connections. This forces researchers to frame their questions and approaches in alignment with these terms, otherwise the research does not get funded. Ethnographic research funded through these mechanisms is equally implicated in generalisation. When the same participants get asked similar questions by researchers over and over, a particular narrative becomes perpetuated. In this way, both qualitative and quantitative forms of evidence are used as instruments of power.

A.D.: In the field of low-carbon infrastructure (often wrongly called renewable energy) research[15], it is possible to make a choice about whose voices will be prioritised. Will it be the corporations or the autonomous insurrectionaries and indigenous farmers?[16-18] In the academic research space, questions need to be asked about whether knowledge should be marketised or devoted to protecting habitats, the waters, the trees, and the soil, and to reveal what is hidden by extractive companies which typically use tactics used to divide and conquer communities through social development, green rhetoric, advertising, or violent repressive strategies.[19,20] Research should generate the best kind of information and knowledge so as to uncover what Power is trying to hide. Researchers should ask whether knowledge production is the best way to support ecologies and people. We need to question whether speaking truth to Power actually works. Arguably, Power is well aware of what it is doing, what its objectives are, and is not particularly concerned with the harm that it causes.[21] Critical research should be focused on actually speaking to people to understand environmental conflicts, rather than just generating new working groups or bureaucracies to continue the existing token procedural adjustments. If research in low-carbon infrastructures and wind and solar generation (more accurate terms than renewable energy) only creates new governmental standards, committees, and checkboxes, is real change being achieved or simply a change in procedures and definitions? Evidence in the solar/wind generation and energy studies can also be used by different kinds of hydrocarbon sectors in an attempt to support continued extraction, which muddies debates, policies, supply chains, and accounting procedures.[8,15] This happens while not a single government is reflecting on past failures and actively working to understand the problems.[22] Instead, governments are failing to take environmental or climate policy seriously.

T.M-W.: Power is often used to manipulate. In the scholarly publication space, evidence shows that the scholarly publication landscape benefits from, amplifies, and showcases topics and questions that relate mostly to the majoritarian view of the world, which is hegemonic, i.e. white male. A key study on this matched author identity to topic and discipline.[23] It found that an author's race and gender identity map very closely to the topics that they study, and, even when controlling for discipline, with authors in the same discipline writing on the same topics with the same keywords, black women would be cited less on average than white men. One way that power manifests is in the identity profile of publications in scholarly literature. The "eugenics map"[23] shows how existing social hierarchies of race and gender, invented and imposed by a very narrow segment of the population to highlight and amplify the position of a very narrow segment of the population, namely, white men and white women, also show up in whose research gets published and cited the most. The analysis showed that publications data was very stratified by race, with, always, men first, and women second, within each of the race categorisations. This social power hierarchy manifests in publications in terms of citation counts and recognition of who is the owner of the knowledge, and who is attributed to being an expert in a field. The study was expanded to look at institutional affiliation, to see how race and gender intersected with being at a prestigious institution. The findings held that the ranking of research is lower when it is published by black women, even when they are affiliated with higher ranked institutions. These findings show that, in the terrain of knowledge production, power manifests in a very structural way and that structure is compounded over time. These inequities are systemic, and therefore require change at leadership and policy levels.

Evidence as resistance to injustice

In the third round, we asked how creating and sharing evidence can be an act of resistance to injustice.

L.A.: The case of schizophrenia research is instructive. This is a very underfunded area, despite the prevalence of the condition in low- and middle-income countries. There is some very exciting research being produced in this field, such as intersectional collaborations that are empowering individuals and communities to be active agents in generating knowledge and shaping research agendas.[24] This is challenging with a condition like schizophrenia; people living with it often have challenges to their capacities (that is the nature of the illness). There is a movement happening to involve people living with schizophrenia in research on the topic, which is also causing researchers to take a step back and question academic and intellectual privilege. Debates about evidence and power can be challenging to follow for someone who does not have educational or mental health privilege, and research on the illness can be inaccessible to those who have it. Recently, the US National Institutes of Health (NIH), one of the biggest funders of medical research internationally, awarded a pinnacle research (R01) grant of USD3 million to a research team, including, as co-investigator, a person with lived experience of schizophrenia. Ten years ago, this would have been inconceivable. This is cause for hope, and an opportunity to begin to undo some of the biases inherent in the practice of science.

An interesting study analysed PhD papers that were produced over 10 years in America across disciplines.[25] It found that the more underrepresented a person was in terms of gender or race in their discipline, the more likely they were to introduce novel concepts. So, in the early 1980s, a woman in computer science was more likely than a man to produce something novel in their research. A man in nutrition was more likely than a woman to produce something novel, non-white people in sociology were more likely to produce something novel, while a man in psychiatry was less likely to produce something novel. This proves that diversity can bring novelty to science, which means better science. Although the same study also found that this novelty does not necessarily translate to impact and implementation, it still offers hope in that it shows the massive value in diversity. In psychiatry, by partnering with people with lived experience of the illness we research, and bringing their diverse experiences into research design, we can create better science. Cautious optimism about the way forward is warranted.

L.M.: Something that is discussed a lot in epistemology and philosophy is the idea that seeing things from different perspectives in no way voids the effort to achieve objectivity, but actually strives towards more possibility of objectivity, because you start to correct for the limitations of those who see things only through a very narrow perspective based

on their own cultural privileges, or racial privileges or class privileges of the past.[26,27] This is different to an argument about the absence of objectivity; it is actually striving towards a better definition of it.

Nevertheless, in the social sciences, there can sometimes be a duplicity towards diversity that results in the exact opposite, even if a more plural field is created. When diversity is prioritised in recruitment processes, on paper, the candidates might come from the Global South, but they all might have studied at elite institutions in the West. In practice, then, the so-called effort towards diversity can lead to class biases that privilege people from poorer regions who are at the top of the class hierarchies in their own nations. Although it is crucial to validate the importance of people in the academy from different racialised backgrounds, we also need to consider other forms of privilege and bias that might come into play. Although academic research and knowledge production should continue to address and speak truth to Power, we should also remember that Power does not care. This forces researchers to be bold about certain achievements; for example, studying at an elite institution, when it comes to an effort towards plurality, will have unintended effects. An honest conversation is needed about those effects without that honesty becoming an apology for the status quo. The status quo needs to change, but it will change in imperfect ways. That leads to a necessity for compromise — a word that needs to come to the forefront as a new radicalism.

Compromise *is* a type of radicalism, especially in a time when the political spectrum is divided by proclamations of achieving a certain type of purity on one side or the other. On the Left, there is pressure to speak one way on every issue or else face being outcast. On the Right, there is growing tolerance for extremism and hate politics. These are both problems. Compromise should be validated as a radical rather than defeatist solution. Optimistic language is the way to go to create political change. I think the Left has become overly reliant on a language of negativity. A language of emancipation gives hope.

A.K.: Who gets validated as a knowledge holder and is able to contribute to discourse in particular ways is central to these debates. Ethnography conducted by and with Pacific researchers and collectives is used to advocate on behalf of Indigenous Pacific peoples and to intervene in climate policy and gender policy.[28] In the Pacific, as in many regions of the world, limited research has been done into climatic and environmental impacts on transgender and [other] queer people. A transgender and queer group in Fiji, called DIVA for Equality, has been conducting "ethnographic" work within their community for several years to produce policy reports and publicly available information around the kinds of discrimination and oppression that transgender but also [other] queer people are experiencing in terms of access to resources, such as shelters during disasters, and materials for rebuilding. Similarly, PANG Fiji produces research reports on economic neo-liberalism and resource extraction in the region, including environmental impacts.[29] These examples show how independent community organisations rely on long-term relationships, kinship networks, and trust. This allows them to conduct their own research and set the agenda. Communities experiencing ecocide and climatic impacts gather evidence in their own voices and for their own purposes. This fundamentally destabilises the idea that the university, think tanks, and governments are where all knowledge is produced.

A.D.: A key question in this debate is what constitutes diversity in knowledge. In elite universities, even though there might be diversity in gender and race in the classroom, it sometimes feels like everyone is from the same gated community. A lot of research in the field of low-carbon extraction simply creates new bureaucracies. In military manuals, bureaucratic processes are noted as a way to sustain occupation.[30] This is simply another version of neo-colonialism. The term ecocide could also potentially be used in problematic ways; for example, the way LGBTQIA+ rights were used to justify overseas interventions by Hillary Clinton.[31] Researchers should, ideally, do everything in their power to produce knowledge that can help people defend their territories and defend their habitats. Evidence should be open access; researchers should be encouraged to do outreach with their findings. From this perspective, universities themselves are the obstruction to social change. Instead of being sustained as spaces that facilitate deep ideas and conversations, they have turned into competitive bureaucracies competing for students and grants. Regardless, the central existential question for researchers should be: can we place ourselves outside of the conflicts that we study? Are we looking "through the window"[32], placing living processes and existences in Petri dishes to poke at and analyse, or do we remember that we are also in the Petri dish?

Researchers should strive to create and widen space inside and outside universities, and to really make social change. This requires thinking seriously about social change and diversity. What are the politics we are trying to create? These definitions matter. Evidence should help to improve the quality of relationships amongst people, with trees, with rivers, with our surroundings, so collectively we can move towards a society that is truly socially and ecologically renewable. When researchers act from a place of just trying to hold onto a job, criticality and integrity are compromised. At this juncture in human existence, evidence is about more than resistance, it is about survival.

T.M-W.: Evidence-based resistance can produce serious, real change for people in their lives. From this perspective, research is never in vain, and evidence always offers hope. In the quantitative world, there is hesitancy to discuss positionality, critical approaches, and critical lenses, but that is changing. New work is addressing the underlying structures that shape the meaning of data and how they are used. There is growing acknowledgement of how power structures envelop all aspects of life, including science and theory. This is creating opportunities to carve out new spaces and experiment with marrying very quantitative, big data, and AI approaches with and critical approaches. This can be summarised as "emancipatory data science"[33]. In this framing, the voices of the marginalised are centred, as well as their experiences and logics. This is about more than getting more people from these marginalised communities involved in research alongside an elitist few. It is about every researcher being able to consider their background, their origins, their privilege, and to integrate that into their knowledge production. Participatory research that centres data from and voices of those most marginalised allows for theorising emancipatory perspectives on the world. This is impossible without partnership and collaboration with the community. Researchers, even in positions of modest to moderate power, can create spaces to integrate and amplify good research. These kinds of acts of resistance are happening everywhere, even in academia.

Consider the example of Howard University, in the USA, a historically black college in Washington, DC. I made a conscious choice to leave a highly prestigious, predominantly white institution in favour of Howard, because during my time at the previous institution, I was in a constant and exhausting state of intellectual resistance and [received] overtly racist remarks from teachers and hostile behaviour from students. The shift to a black institution meant that significant time was saved from having to protest, and instead I could simply focus on my studies. Relatively modest acts of resistance, such as rejecting the doctrine that predominantly white institutions are best, or citing marginalised, minoritised writers, matter. This means that researchers must actively look for research and scholarly publications from structurally underrepresented groups. As well as this, researchers should choose as much as possible to publish open access. Even those who have not yet gained tenure can consider these routes, so as to be their full selves in scholarly spaces.

Conclusion

A.F. and M.I.: This conversation has illuminated questions about how we share and archive practices of 'speaking evidence to power'. Evidence does not "speak for itself" and usually needs to be explicated or interpreted by scientific actors, which in turn means that it can be appropriated or deployed with specific agendas, either by those in power, or those resisting authority. The emancipatory possibilities attached to evidence will therefore shift depending on context. Evidence means something different in each disciplinary research setting, is shaped by specific ideological leanings, and is deployed in different ways depending on political agendas. There remains the problem that "Power doesn't care about evidence". Indeed, power will always find ways to twist evidence to its own purposes, and to serve its own agendas. What role then, could evidence continue to play in empowering those who are exploited or marginalised?

There would be benefit to sharing more examples of studies and research projects, whether inside or outside of academia, that are doing really innovative things, combining quantitative and qualitative methods and participatory research of some kind, whether that is through the ways it is done in medical research, through a kind of co-production/co-creation, or through working with community organisations.

Secondly, there continues to be a tension to navigate between knowledge production inside and outside of academia. Our seminar speakers each explored specific tensions in the projects of producing university- and government-validated knowledge and producing counter-knowledges with communities. Whether it is pressures produced by funding applications, promotion processes, seeking job stability or higher levels of influence, work inside academia will always have to conform to certain frameworks and standards of what counts as knowledge and evidence. Even within the same disciplinary boundaries, evidence can sometimes have many faces, or wear different masks. Huge epistemological debates take place within disciplines as well as between them, and as such we need to continue to explicitly engage in debates about what kinds of data are considered convincing and why, in which contexts.

Although there is huge potential in a politics that seeks to amplify the co-production of knowledge outside of the university, this has in some ways become more complex in the past decade as universities have taken increased interest in public engagement, and research impact and evidence (or some kinds of evidence) has come to exert more power. This raises questions about whether academia and its pressures for public engagement is the right place for activism and change-making.

This interdisciplinary conversation has dived deep into power and how it operates through knowledge construction and knowledge making. It has shown how all forms of science are structures of communication, are practices of communication, and are all completely embedded both within structures and histories of power as well as different modalities of defining what evidence is. These questions, and ongoing conversations in response, should remain at the heart of all kinds of ontologies of what research knowledge creation is and can be.

Acknowledgements

We thank Fumani Jwara for logistical support in organising the Evidence & Power seminar series, and Nakishka Skriker for editorial support in preparing this Structured Conversation.

Declarations

We have no competing interests to declare. We have no AI or LLM use to declare. All authors read and approved the final manuscript.

References

1. Scott JC. Seeing like a state: How certain schemes to improve the human condition have failed. New Haven, CT: Yale University Press; 2020. https://doi.org/10.12987/9780300252989
2. Bryant-Tokalau J. Indigenous Pacific approaches to climate change. Dunedin: Palgrave Pivot Cham; 2018. https://doi.org/10.1007/978-3-319-78399-4
3. Teaiwa KM. Consuming Ocean Island: Stories of people and phosphate from Banaba. Bloomington, IN: Indiana University Press; 2014.
4. Kanngieser A. Weaponizing ecocide: Nauru, offshore incarceration, and environmental crisis. Contemp Pac. 2020;32(2):492–502. https://doi.org/10.1353/cp.2020.0042
5. Kanngieser A, Tau J. Mining the deep [webpage on the Internet]. c2019 [cited 2023 Jun 15]. Available from: https://amkanngieser.com/posts/mining-the-deep
6. Foucault M. The order of things: An archaeology of the human sciences. New York: Vintage; 1994.
7. Sullivan S. What's ontology got to do with it? On nature and knowledge in a political ecology of the 'green economy'. J Political Ecol. 2017;24(1):217–242. https://doi.org/10.2458/v24i1.20802
8. Dunlap A. The green economy as counterinsurgency, or the ontological power affirming permanent ecological catastrophe. Environ Sci Policy. 2023;139:39–50. https://doi.org/10.1016/j.envsci.2022.10.008
9. Robbins P. Political ecology: A critical introduction. Malden, MA: Blackwell Publishing; 2004.
10. Benjaminsen TA, Svarstad H. Political ecology: A critical engagement with global environmental issues. Cham: Palgrave Macmillan; 2021. https://doi.org/10.1007/978-3-030-56036-2
11. Plancikova D, Duric P, O'May F. High-income countries remain overrepresented in highly ranked public health journals: A descriptive analysis of research settings and authorship affiliations. Crit Public Health. 2021;31(4):487–493. https://doi.org/10.1080/09581596.2020.1722313
12. Harris M, Marti J, Watt H, Bhatti Y, Macinko J, Darzi AW. Explicit bias toward high-income-country research: A randomized, blinded, crossover experiment of English clinicians. Health Aff. 2017;36(11):1997–2004. https://doi.org/10.1377/hlthaff.2017.0773
13. McGoey L. No such thing as a free gift: The Gates Foundation and the price of philanthropy. London: Verso Books; 2015.
14. McCoy D, Kembhavi G, Patel J, Luintel A. The Bill & Melinda Gates Foundation's grant-making programme for global health. Lancet. 2009;373(9675):1645–1653. https://doi.org/10.1016/S0140-6736(09)60571-7
15. Dunlap A. Does renewable energy exist? Fossil fuel + technologies and the search for renewable energy. In: Batel S, Rudolph D, editors. A critical approach to the social acceptance of renewable energy infrastructures: Going beyond green growth and sustainability. Cham: Springer International Publishing; 2021. p. 83–102. https://doi.org/10.1007/978-3-030-73699-6_5
16. Dunlap A. Insurrection for land, sea and dignity: Resistance and autonomy against wind energy in Álvaro Obregón, Mexico. J Political Ecol. 2018;25(1):120–143. https://doi.org/10.2458/v25i1.22863
17. Dunlap A. 'Agro sí, mina NO!' the Tía Maria copper mine, state terrorism and social war by every means in the Tambo Valley, Peru. Political Geogr. 2019;71:10–25. https://doi.org/10.1016/j.polgeo.2019.02.001
18. Dunlap A. Bureaucratic land grabbing for infrastructural colonization: Renewable energy, L'Amassada, and resistance in southern France. Hum Geogr. 2020;13(2):109–126. https://doi.org/10.1177/19427786209180
19. Verweijen J, Dunlap A. The evolving techniques of the social engineering of extraction: Introducing political (re)actions 'from above' in large-scale mining and energy projects. Political Geogr. 2021;88, Art. #102342. https://doi.org/10.1016/j.polgeo.2021.102342
20. Granovsky-Larsen S, Santos L. From the war on terror to a war on territory: Corporate counterinsurgency at the Escobal mine and the Dakota Access Pipeline. Can J Latin Am Caribbean Stud. 2021;46(1):121–145. https://doi.org/10.1080/08263663.2021.1855892
21. Churchill W. On the necessity of armed struggle: Reflection on the RAF and the question of moving forward. In: Moncourt A, Smith J, editors. The Red Army Faction: A documentary history. Montreal / Oakland, CA: Kersplebedeb Publishing / PM Press; 2009.
22. Stoddard I, Anderson K, Capstick S, Carton W, Depledge J, Facer K, et al. Three decades of climate mitigation: Why haven't we bent the global emissions curve? Annu Rev Environ Resour. 2021;46(1):653–689. https://doi.org/10.1146/annurev-environ-012220-011104
23. Kozlowski D, Larivière V, Sugimoto CR, Monroe-White T. Intersectional inequalities in science. Proc Natl Acad Sci USA. 2022;119(2), e2113067119. https://doi.org/10.1073/pnas.2113067119
24. Morgan J. Nev Jones: Fixing a broken system. Lancet Psychiatry. 2022;9(1):18. https://doi.org/10.1016/S2215-0366(21)00470-3
25. Hofstra B, Kulkarni VV, Munoz-Najar Galvez S, He B, Jurafsky D, McFarland DA. The diversity–innovation paradox in science. Proc Natl Acad Sci USA. 2020;117(17):9284–9291. https://doi.org/10.1073/pnas.1915378117
26. Harding SG. Objectivity and diversity: Another logic of scientific research. Chicago, IL: The University of Chicago Press; 2015. https://doi.org/10.7208/9780226241531
27. Sullivan S, Tuana N, editors. Race and epistemologies of ignorance. Albany, NY: State University of New York Press; 2007. https://doi.org/10.1353/book5200
28. Diverse Voices and Action for Equality. Unjust, unequal, unstoppable: Fiji lesbians, bisexual women, transmen and gender non confirming people tipping the scales toward justice [document on the Internet]. c2019 [cited 2023 Jun 15]. Available from: https://divafiji.org/wp-content/uploads/2020/12/DIVA-RESEARCH-REPORT-2019.pdf

29. Blue Ocean Law, The Pacific Network on Globalisation. Resource roulette: How deep sea mining and inadequate regulatory frameworks imperil the pacific and its peoples [document on the Internet]. c2016 [cited 2023 Jun 15]. Available from: https://ramumine.files.wordpress.com/2016/06/resource-roulette-how-deep-sea-mining-and-inadequate-regulatory-frameworks-imperil-the-pacific-and-its-people.pdf

30. Dunlap A, Arce MC. 'Murderous energy' in Oaxaca, Mexico: Wind factories, territorial struggle and social warfare. J Peasant Stud. 2022;49(2):455–480. https://doi.org/10.1080/03066150.2020.1862090

31. U.S. mission to international organizations in Geneva. Secretary of state Clinton Human Rights Day speech – free and equal in dignity and rights [webpage on the Internet]. c2011 [cited 2023 Jun 25]. Available from: https://geneva.usmission.gov/2011/12/06/free-and-equal/

32. Gelderloos P, Dunlap A. 'The poisons are already in here with us': Framing for ecological revolutions from below. Globalizations. 2023;21:1–18. https://doi.org/10.1080/14747731.2023.2225306

33. Monroe-White T. Emancipatory data science: A liberatory framework for mitigating data harms and fostering social transformation. In: Proceedings of the 2021 Computers and People Research Conference. New York: Association for Computing Machinery; 2021. p. 23–30. https://doi.org/10.1145/3458026.3462161

AUTHORS:
Philippa Kerr[1]
Kevin Durrheim[1]
Maria Schuld[1]
Davide Morselli[2]

AFFILIATIONS:
[1]Department of Psychology, University of Johannesburg, Johannesburg, South Africa
[2]Swiss Centre of Expertise in Life Course Research (LIVES), University of Lausanne, Lausanne, Switzerland

CORRESPONDENCE TO:
Kevin Durrheim

EMAIL:
kevind@uj.ac.za

DATES:
Received: 08 Dec. 2023
Revised: 22 July 2024
Accepted: 07 Aug. 2024
Published: 04 Dec. 2024

HOW TO CITE:
Kerr P, Durrheim K, Schuld M, Morselli D. A thematic analysis of South African opinions about COVID-19 vaccination on Twitter. S Afr J Sci. 2024;120(11/12), Art. #17423. https://doi.org/10.17159/sajs.2024/17423

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

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

EDITOR:
Floretta Boonzaier

KEYWORDS:
COVID-19, vaccination, Twitter, pro-vaccination, anti-vaccination

FUNDING:
South African National Research Foundation (137755), South African Centre for Digital Language Resources (OR-AAALV), Swiss National Sciences Foundation (10001AL_205032)

A thematic analysis of South African opinions about COVID-19 vaccination on Twitter

Vaccine hesitancy is a public health concern in South Africa and internationally. Literature on vaccine hesitancy associates this with mistrust of the government. We present a qualitative analysis of opinions about COVID-19 vaccination expressed by South African Twitter (now X) users during the first year of the vaccine rollout in South Africa. We conducted a thematic analysis of 800 randomly selected tweets containing vaccine-related keywords, sampled from four time periods in 2021. We categorised comprehensible South African non-news tweets as pro-vaccination (24.75% of sample), anti-vaccination (20.25%) or ambivalent (4.5%), and then identified themes. Among pro-vaccination tweets, the most common themes were criticism of the government's handling of vaccine procurement and the rollout; concerns that the vaccine was urgently needed and/or not being made available fast enough; and statements that vaccines were safe and/or effective against COVID-19. Among anti-vaccination tweets, the most common themes were claims that the vaccine was harmful or too risky; suspicion of the government's intentions with respect to the vaccine it was offering the public; and opposition to mandatory or 'forced' vaccination. Criticism and mistrust of the government were present among both pro- and anti-vaccination tweets, though for different reasons. We discuss this in light of literature recommending trust-building as a response to vaccine hesitancy.

Significance:

Numerous studies recognise mistrust in the government as a correlate of anti-vaccination opinions, but our findings suggest that holders of both pro- and anti-vaccination opinions in South Africa mistrust the government – albeit for different reasons. Several South African authors propose 'trust-building' as a solution to vaccine hesitancy and refusal, but we suggest that in a context of government corruption, it is not more trust that the South African public needs, but more critical literacy in order to discern when the government is and is not acting in the public (health) interest.

Introduction

Although COVID-19 became an opportunity for a resurgence of the anti-vaccination movement, opposition to vaccinations is not new, and in fact is almost as old as vaccination itself.[1] Several years before the appearance of COVID-19, vaccine hesitancy and refusal had been identified by the World Health Organization (WHO) as a threat to global health[2,3]; these have also been factors limiting the uptake of vaccination for COVID-19[4,5]. A summary of findings from South African surveys conducted in 2020 on willingness to get vaccinated for COVID-19[4] suggests that only between roughly half and 80% of South African adults were willing to get vaccinated if and when a vaccine became available and that vaccine hesitancy actually increased over the course of the pandemic. The latest Africa Center for Disease Control and Prevention statistics suggest that approximately 40% of the South African population has been fully vaccinated for COVID-19.[6]

The WHO identifies five common themes in what it calls 'vaccine denial' beliefs.[2] The first, 'threat of disease', refers to arguments that vaccine-preventable diseases offer no significant threat, either because they no longer exist or because they are not serious. Thus, according to this argument, there is no need for vaccination. The second topic, 'trust', refers to (a lack of) trust in the medical, scientific and government institutions promoting vaccination. To the extent that information coming from these sources is viewed with scepticism, simply offering or repeating factually accurate information about vaccines is unlikely to challenge the misinformation that is circulated by the anti-vaccination movement. The third theme, 'alternatives', refers to alternative remedies that are claimed to be safer and/or more effective than vaccines. The fourth and fifth themes, 'effectiveness' and 'safety', refer to concerns that vaccines are not effective in preventing disease and are in fact unsafe, with harmful side effects, including illness and death.

The rhetoric of anti-vaccination discourse

There is a growing body of literature confirming and expanding on this list of themes in anti-vaccination beliefs, including from South Africa since COVID-19.[4,7-13] In this literature, anti-vaccination opinions are almost universally associated with mistrust in the government and other forms of authority.[1,11] Some of this work goes further than listing common beliefs and themes, however, and illustrates the reflexive, socially embedded and rhetorically sophisticated nature of anti-vaccination discourse. Rozbroj et al.[11] surveyed members of the public who self-identified as part of the anti-vax movement in Australia, asking them open-ended questions about what being part of this movement meant to them. Respondents persuasively presented themselves as pro-science, pro-choice, responsible parents, brave, enlightened, critical thinkers, highly informed and highly engaged with the health system. They took issue with the label 'anti-vaccination', seeing it as a derogatory term and as evidence of how they were stigmatised, which strengthened their identity as a movement of courageous people speaking 'inconvenient truths'. The authors argued that the anti-vaccination movement 'is strengthened by hostility towards it, defining itself as bravely fighting for an important cause in the face of undue hardship' (p. 2). Hence, anti-vaccination advocates do not merely provide alternative (mis)information: they show clear awareness of the evidence for the effectiveness of vaccines ('the mainstream'), and their arguments are geared towards rebutting

or undermining this. Hence, anti-vaccination beliefs cannot be attributed simply to a lack of correct information, or solved with the provision of more correct information.[1,11] In contrast to the volume of work on anti-vaccination discourse, there is relatively little work on pro-vaccination opinions (but see Connoway et al.'s analysis of pro- and anti-vaccination South African Facebook pages[7]).

Trust and the politics of vaccination opinions

In South Africa, no major political party took an anti-vaccination position during the COVID-19 pandemic[4], and public opinion about vaccination has not been overtly party-politicised and polarised in the same way as, for example, in the United States of America[14], although Cooper et al. did find that support for the African National Congress (ANC) was associated with greater willingness to vaccinate than support for other parties[4]. Nevertheless, again, one of the most common themes in the South African vaccine hesitancy literature is mistrust of the government, and the 'trust-building measures' that are consequently needed to counter it.[7,8,13] For example, Gittings and colleagues[8] have argued that "mistrust in government and international health systems has emerged as a concerning determinant of vaccine hesitancy and one that is particularly important to tackle" (p. 301). Similarly, Cooper and colleagues[4] argue that building people's confidence in COVID-19 vaccines . . . needs to form part of broader development and trust-building measures that focus on relationships, transparency, participation, and justice. For example, strong leadership and clarity around responses to the COVID-19 pandemic, including . . . vaccines, is important. (p. 930)

Although we also found numerous expressions of mistrust in the government among anti-vax tweets (one of the most frequent anti-vaccination themes we called 'mistrust of the government, therefore mistrust of the vaccine'), these data provide grounds for questioning whether building trust in the government is the straightforward solution to vaccine hesitancy. Firstly, pro-vax tweets were not especially trusting of the government. There were many pro-vax concerns about whether the government could be relied on to deliver vaccines efficiently, timeously and without corruption. Secondly, the call for trust-building introduces a conundrum where the same institutions that people already mistrust must take on the job of convincing people to trust them, thus overlooking the possibility that mistrust about vaccines can take on a life of its own and be unrelated to whether the government is actually acting in the public interest.[1] In the following section, we describe our research questions and methods and then present findings from our thematic analysis of South African tweets related to vaccination for COVID-19.

Research methods

Research questions

Our research questions were: what was the content of vaccination-related opinion on South African Twitter (now X) during the COVID-19 vaccination rollout? Did this vary over the course of 2021? To what extent did these opinions contain similar themes to those already identified in the vaccine hesitancy and refusal literature?

Data collection and sampling

We extracted 574 197 original tweets from the Twitter API, the full sample identified as vaccine-related tweets from South African Twitter users in 2021. Vaccine keywords included: vaccine, vaccination, vax, antivax, anti-vax, anti-vaccine, antivaccine, vaxed, Vaxxed, unvaxed, unvaxxed, vaccinated. The geographic keywords were chosen using a multi-step procedure, in which we geolocated South African Twitter users, extract their followers and extract keywords (words, hashtags, ngrams, mentions) from their vaccination-related tweets that are (a) used frequently and (b) clearly reference a South African context.

On the basis of preliminary content analysis and a prior understanding of the history of the pandemic, we divided the tweets into four phases:

1. 01/02/2021 – 20/02/2021: Announcement of vaccination programme ($N = 129\,327$).
2. 21/02/2021 – 20/06/2021: Low-frequency tweet period, in between waves; front-line health workers and people over 60 are being vaccinated ($N = 104\,689$).
3. 21/06/2021 – 23/11/2021: The mid-year Delta wave; lockdown restrictions return; progressively more age and employment categories are being vaccinated (This is also the first wave after vaccination began, raising doubts over the effectiveness of vaccines.) ($N = 219\,074$).
4. 24/11/2021 – 30/12/2021: The Omicron variant period, discovered on 23 November ($N = 121\,107$).

We then randomly selected 1000 keyword-containing tweets from each phase. These tweets were recorded in four .txt files, with one tweet per line. In the process, the tweets were removed from their threads, and all tweeters' identities were removed, so the tweets are anonymised, but any other users' handles the tweeter tagged were retained. For anonymity's sake, we have manually removed handles other than those of politicians, political parties and news organisations. Pre-processing also removed all punctuation, special characters such as '&' and '%' (but not #), images and videos from all tweets, and it made capital letters lower case. The absence of punctuation and capital letters sometimes made it difficult to understand the meaning of tweets, especially those with numbers that probably originally contained decimals or percentages. Emojis were retained and are reproduced below.

Qualitative and quantitative analysis of tweets

The analytic process was open-ended and inductive: we had no initial hypotheses about what we would find, and we did the coding before we had read the vaccination opinion literature extensively. This means that these findings can be seen as independent of those from other studies and can be usefully compared with them, an important way of establishing the credibility of these findings. We used a combination of thematic analysis and content analysis, which was appropriate for our aim of capturing the quality and quantity of vaccination opinions over the course of the rollout.[15]

Our analysis proceeded in three steps. We began by reading through the four sets of 1000 tweets, and making notes on the topics and styles in them. We then created a set of coding spreadsheets. The first 200 tweets of each text file were selected for coding: these were pasted into the first columns of four MS Excel spreadsheets respectively. In the next column, each tweet was then assigned a number from 1 to 7, thus: 1 for tweets that we read as being pro-vaccination, 2 for anti-vaccination, 3 for ambivalent about vaccination (expressing both pro- and anti-vaccination views), 4 for tweets that offered no clear opinion about vaccination (for example, asking questions about vaccination but offering no own position), 5 for tweets that appeared to come from outside South Africa (even if they mentioned South Africa), 6 for international and local news tweets and official statements and 7 for tweets that were incomprehensible and/or not about COVID-19 vaccination. This was the first step.

In the second step, tweets coded as pro-vaccination, anti-vaccination or ambivalent (i.e. opinion tweets) were then further assigned as many thematic tags as necessary to capture their contents. In the process of reading the 800 tweets from the four phases, 16 pro-vaccine themes were identified, 2 that were only associated with ambivalent tweets, 3 themes that were not restricted to any particular position and 17 anti-vaccination themes. Later, some themes were merged together, and others were split into more specific themes. We report the results of this qualitative analysis in Supplementary table 1, where we describe and illustrate each of the emergent themes of pro- and anti-vax tweeting. (For brevity's sake, we have omitted ambivalent themes from the table.)

Finally, in the third step, the number of tweets referencing each theme was counted, resulting in the 'frequency' count in Supplementary table 1. Quantifying our themes in this way allowed us to track changes in the expression of different vaccination opinions over time.

Trustworthiness in the analysis

On what basis did we assign tweets to categories? In qualitative analysis, the analyst's job is to read for meaning. We were not working with an already-existing framework that distinguishes pro- from anti-vax opinions; we were creating the framework as we went along, based on an ever-expanding understanding of the parameters of the Twitter vaccination debate. There is an inherent degree of ambiguity in all discourse[16], making a search for total clarity inappropriate; however, in

this data set, some tweets were more ambiguous than others. Consider the following tweet, which contains a clear pro-vaccination stance:

> *the nurse was so quick i didnt even get to take a pic with the needle in my arm* 😩 *im taking* [redacted name] *with next time* 😂 *annyyywayyyy your favourite aunty is halfway there* 💉🥳🙌🏽 *#vaccinerolloutsa #firstdose #35to49gang*

As we came to discover, this tweet, in which the tweeter implies she has had the first of two shots, was part of a genre of pro-vax tweets announcing that the tweeters themselves had been vaccinated. However, it was sometimes less clear whether a tweeter's position was pro- or anti-vax, for example:

> *@news24* [names redacted] *they did actually in countries like france you cant even sit in a restaurant without proof of vaccination hence there were protests but people have accepted they even have qr code scanners so they can check on the spot when were you vaccinated and which vaccine did you take*

This concern with needing to show proof of vaccination to access services was often associated with anti-vax opinion in this data set, but in this case, the tweeter does not ultimately offer their own opinion. Hence, we coded this tweet as '4' – not containing a clear opinion. We recognise that some might disagree, arguing that this is definitely an anti-vax tweet. We spent a long time mulling over ambiguous tweets, and hence, we do not claim that our scheme is the only possible way of representing the range of opinions in this data set. However, we do claim that the analysis was rigorous and comprehensive, that the themes we identified are confirmed by much of the literature, and that the frequencies reported in Supplementary table 1 provide a broadly representative overview of the data.[17] As Silverman argues, using quantification in qualitative research (that is, counting the number of times each theme appeared in our data set) can lend weight to conclusions if used with a clear sense of what is being counted, and why. Counting these frequencies allows us to avoid 'anecdotalism' by showing how prevalent these vaccination-related concerns were over the course of 2021.

Results

Here we report an analysis of changing themes over time. Figure 1 describes and counts the different kinds of tweets we found in this Twitter sample across the four time phases, and Figures 2 and 3 summarise the most common themes. Exemplary tweets can be found in Supplementary table 1.

Figure 1 shows that South African vaccination talk on Twitter in 2021 was dominated by opinions and news. There were slightly more tweets expressing pro-vaccination opinions in this sample (198 or 24.75% of the sample) than anti-vaccination opinions (162 or 20.25% of the sample), with the proportion of anti-vaccination tweets growing in phases 3 and 4. Ambivalent tweets (36 or 4.5% of the sample) were relatively infrequent. News-like tweets made up nearly a quarter of the sample, being especially prominent in phases 1 and 2.

Figures 2 and 3 show frequencies of the 10 most common themes across the four time phases. We have grouped these into pro-vaccination (Figure 2) and anti-vaccination (Figure 3) themes.

Discussion

We highlight some findings from these figures. First, many of the anti-vaccination themes in this sample correspond with findings from qualitative analyses of vaccine opinions documented elsewhere: in particular, mistrust of the government's intentions with respect to the vaccine it was administering to the public; claims that vaccines are harmful, deadly, untested, risky, ineffective and/or not what they purport to be; the idea that it is really anti-vaxxers who know the truth and are thinking critically for themselves about vaccination; and opposition to 'forced' vaccination on grounds that it violates individuals' rights and bodily autonomy.[1,2,4,7,8,11]

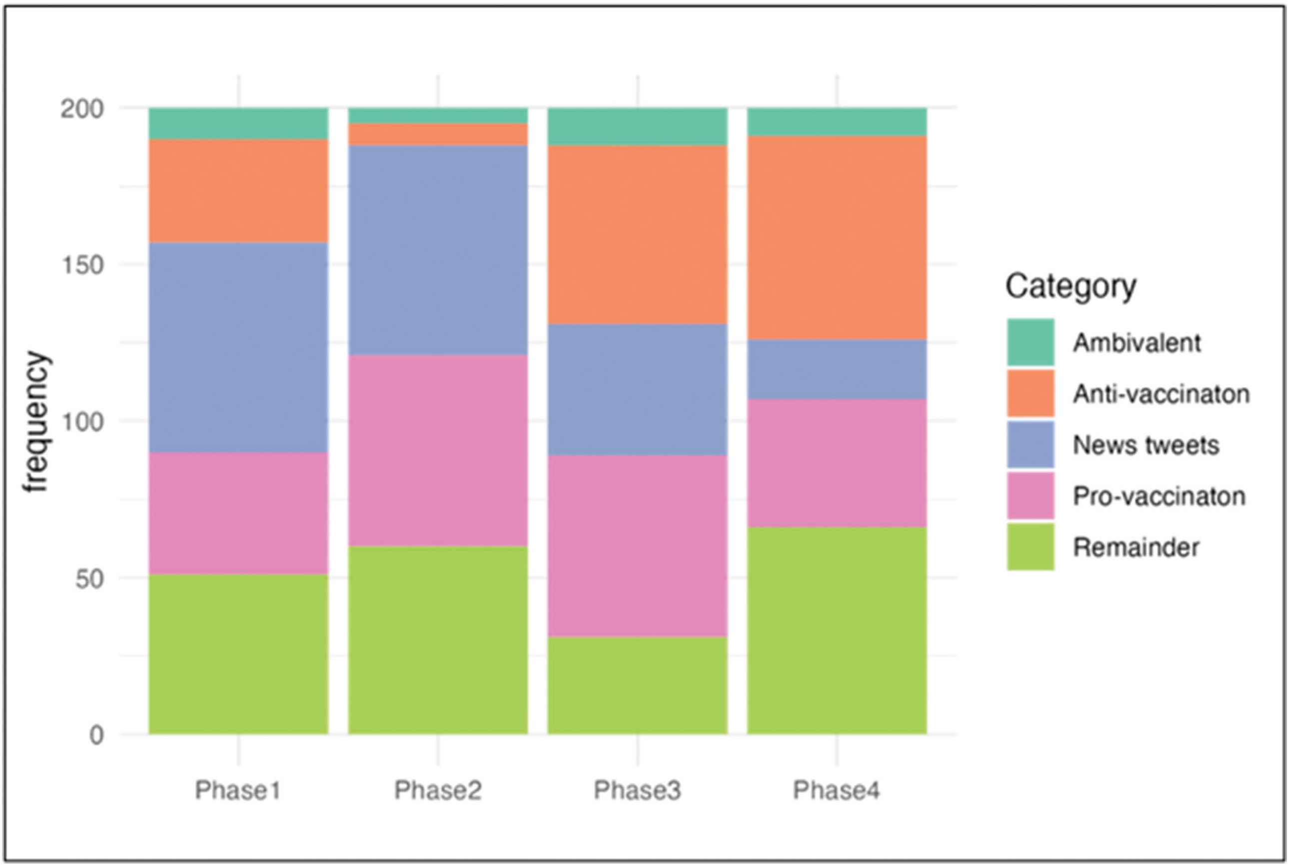


Figure 1: Proportions of tweet types by phase. 'News tweets' include international and local news as well as other news-like reports and official statements of fact. 'Remainder' includes incomprehensible tweets, tweets not about COVID-19 vaccination and tweets probably not from South Africa.

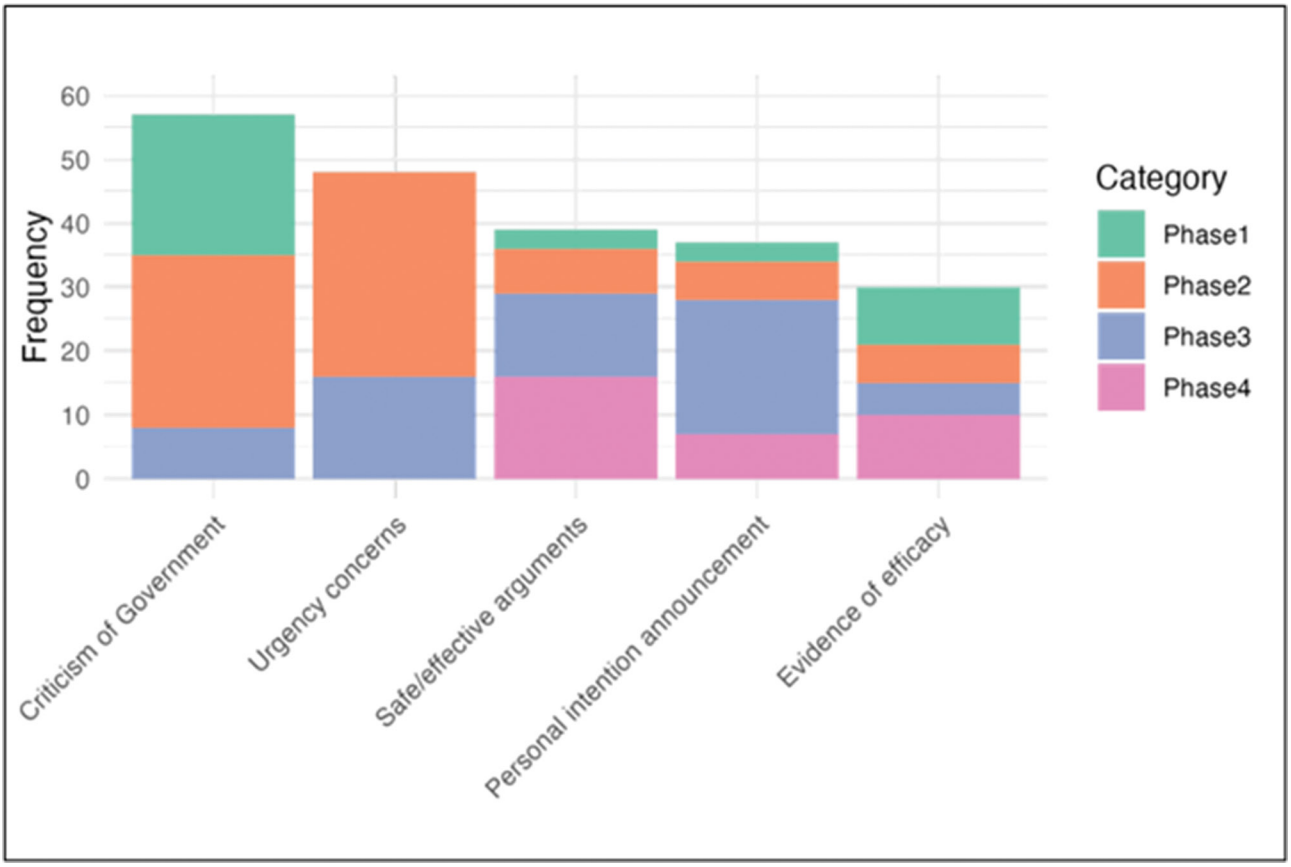


Figure 2: Frequency of top five pro-vaccination themes by phase. Theme 1 – Criticism of government's handling of vaccine procurement and rollout. Theme 2 – Concerns and complaints that vaccination is urgent but not going fast enough. Theme 3 – Statements that vaccines are safe and/or effective against COVID-19. Theme 4 – Announcements of personal intention to get vaccinated, or of having been vaccinated. Theme 5 – Discussions of emerging evidence on vaccine efficacy.

As Connoway and colleagues[7] note, however, "some themes [in anti-vaccination discourse] endure, [while] others are more specific to time and place" (p. 4). Likewise, the WHO acknowledges that although there are some recognisable common themes in vaccine-denialist beliefs, there are also a multitude of different reasons why people hesitate about or refuse vaccination in different parts of the world.[2,3] Supplementary table 1 contains a comprehensive overview of every theme we found in this data set, with an example tweet for each one. A productive direction for future analysis would be to study these in more detail and to ask questions about which aspects of vaccine hesitancy and refusal on South African Twitter are consequences of the international anti-vax movement's influence, and which are more local in origin.

Second, we note that the emergence and disappearance of several pro-vaccination themes seems to correspond clearly to the progress of the rollout in South Africa. Criticism of the government's handling of the process was a relatively common theme in the first half of the year but then declined, presumably as the rollout got underway and the whole adult population became eligible for vaccination. Expressions of urgency that the vaccination rollout was not going fast enough did not appear at the very beginning of the process, but appeared in phase 2 (late February to June), and then also petered out in later phases. Relatedly, personal announcements of having been or intending to get vaccinated peaked in phase 3, starting in July, when eligibility for vaccination dropped from over 60s to include younger age groups (who are also presumably more active on Twitter). Also, approximately a quarter of all tweets were news-like tweets (in the register of a headline) and/or official statements, but these declined proportionately in the last two phases, presumably as news interest in the pandemic tailed off towards the end of the year.

By contrast, anti-vax opinions were relatively uncommon in the first two phases, and in fact almost disappeared in phase 2, but then increased in phases 3 and 4. Concerns with mandatory vaccination, claims about vaccine inefficacy, and claims that COVID-19 was not real or not serious, hardly featured in the first two phases, but then increased significantly in phases 3 and 4. We have no hypothesis to offer explaining the decrease in the proportion of anti-vax tweets in phase 2, but the increase in the last two phases resonates with South African survey findings that vaccine hesitancy actually *increased* as the pandemic and rollout progressed[4], and this seems to have occurred at the same time as news interest was declining.

Third, we note that although much of the literature associates anti-vax opinions with mistrust of the government[4,8,11-13], criticism and mistrust of the government was a strong theme among both pro- and anti-vaccination tweets in this sample. Even though the South African government took a firmly pro-vaccination official stance, pro-vax tweeters were still quick to criticise the government's handling of vaccine procurement and the rollout, latching onto blunders, raising concerns about corruption, and initially complaining that the rollout was going too slowly. And when the rollout eventually got underway, almost no pro-vax tweeters gave credit to government for this. We found only one tweet out of 800 that did so (see Supplementary table 1), compared to 57 tweets that criticised the government's efforts, 37 that announced the tweeter personally had been or would be getting vaccinated and 23 discussing (in a neutral or positive way) logistics and milestones in the vaccination rollout. Otherwise, the complaints dried up towards the end of the year (see Figure 1).

Anti-vax tweets were also critical of the government, but for opposite reasons: a number complained that the government was forcing them to take a harmful vaccine against their will, and/or expressed mistrust of the government's intentions with respect to the vaccine it promoted to the public. Some tweeters explicitly said they had 'trust issues' with the government and needed more reassurance that what they would be getting was a real vaccine. Hence, the difference between pro- and anti-vax tweeters was not that the former trusted and the latter mistrusted

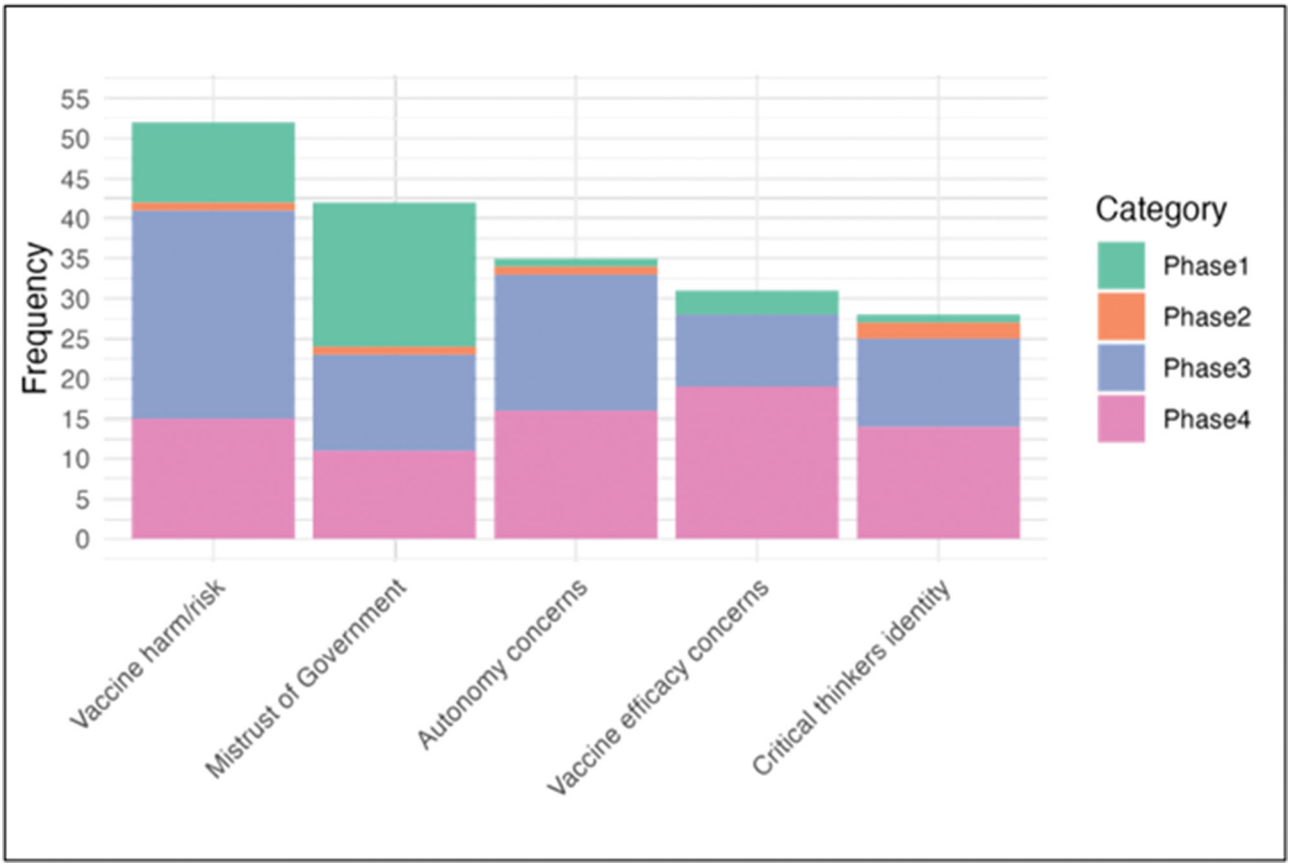


Figure 3: Frequency of top five anti-vaccination themes by phase. Theme 1 – Vaccine is harmful, too risky or insufficiently tested. Theme 2 – Mistrust of the government's intentions w.r.t. the vaccine. Theme 3 – Vaccine mandates violate individual autonomy and rights. Theme 4 – Doubts about vaccine efficacy; claims it doesn't work. Theme 5 – Anti-vaxxers are critical thinkers who can see through the lie.

the government; rather, the difference was in the substance of their mistrust. Pro-vax critics of the government were making an assumption that the (untrustworthy) South African government and the (trustworthy) international vaccine development enterprise were independent of each other, so the government could be held to account for how it handled these vaccines, whereas anti-vaxxers' suspicions tended to assume there was some shady, unclear connection between the government and the origins of the vaccine it purported to be administering, and hence, since the South African government was generally corrupt and not to be trusted, neither was the vaccine.

Conclusions

In thinking about responses to vaccine hesitancy and COVID-19 denialism, it is important to learn lessons from previous studies of the international anti-vaccination movement. Dubé and colleagues note how the anti-vaccination movement works by constantly shifting the goalposts, making the mere provision of accurate information from scientific discoveries an ineffective way of challenging it:

> *It is thus unlikely that accumulation of scientific evidences disproving the causal association between vaccination and different diseases or conditions (e.g., sudden infant death syndrome, autism, diabetes, etc.) will ever stop the anti-vaccination movements. This is well illustrated by the shifting hypothesis linking the measles component of the MMR to autism: once disproved by science, a new hypothesis was generated that focused on additives in vaccines, and then after that, on 'too many, too soon'…Evidence alone does not help reshape these anti-vaccine beliefs.*[1(p.107)]

Moreover, criticising the anti-vaccination movement can have the consequence of strengthening the group's sense of being a persecuted minority bravely speaking the truth.[11] These insights mean that sensitivity and caution are needed in thinking about future strategies to encourage vaccination. Hoare et al. distinguish between hardline anti-vaxxers, who are impervious to persuasion, and those who are initially unsure about vaccination but are open to being persuaded if they find health professionals who will answer their questions carefully and explain the benefits.[18]

We would also be cautious to recommend that the public become 'more trusting' of government, even though we agree that mistrust in the government was a common theme among anti-vax tweets. A degree of mistrust in the government is probably healthy and appropriate, given the South African government's endemic corruption, general indifference to the well-being of its people and past history of AIDS denialism. We should hold our government to account – as many of the pro-vax tweeters in this sample were doing – not mindlessly trust it. The trust-building recommendation also introduces a conundrum where the onus for convincing people to be more trusting is put onto the same entities that people already mistrust. Trying to get around this by saying that information needs to come from sources people do trust[8,12] then potentially creates an endlessly moving target for intervention, and begs the question of why people do not trust in the first place, even when the government *is* acting in the public interest. Here we come back to the crucial insight that the anti-vaccination movement manufactures controversy and mistrust where none previously existed.[1] Anti-vaccination movements exist even in countries with traditionally very high levels of public trust in government, e.g. Norway and Australia. As Cooper and colleagues note, the COVID-19 vaccine rollout in South Africa was *already* a high point of public service delivery[4], yet much of the public rejected it.

Hence, it seems to us that what is needed is greater critical literacy on the part of the public, enabling us to exercise discernment about when the government is acting in the public interest and when it isn't. The 'mainstream media' have already played a crucial role in this regard, as journalists provided the public not only with information about COVID-19 and about vaccination but also about government corruption involving funds set aside for the vaccination programme. Yet the tragedy of vaccine refusal during COVID-19, at least as reflected in this Twitter commentary, was that a large proportion of the public was apparently unable to tell the difference: many tweeters were unable to separate their knowledge of government corruption and service delivery failures from evidence that in this particular case, the government was actually succeeding at doing something in the public interest, for example, the following tweet from our sample:

> *to those who cant wait to take the vaccine if you think the government will ever do anything for your benefit think again #familymeeting #alcoholban*

Hence, we recognise that addressing vaccine hesitancy may imply focus in a different direction from 'more trust': towards public education, critical literacy and patient-led advocacy that holds the government to account and enables people to make their own informed, self-interested health decisions. However, we recognise that if the vaccine-refusing public believes they are already making their own self-interested health decisions and mistrusts the 'mainstream media' as well as the government, then this provides an enormous challenge about how to encourage vaccination uptake in the absence of a societal consensus about what is in fact in the public's best health interests. Although Twitter users are not representative of the whole South African public, social media platforms are becoming increasingly important in spreading opinions and misinformation[19], and their role in promoting the public good requires more research.

Acknowledgements

We thank Martin Mafunda for making the bar graphs.

Funding

K.D.: National Research Foundation of South Africa (Grant UID: 137755) and the South African Centre for Digital Language Resources (SADiLaR) (Grant OR-AAALV). D.M.: People-Opinion Networks: A study of polarisation in word embeddings and social networks in Switzerland and Southern Africa, Swiss National Sciences Foundation, ref: 10001AL_205032.

Data availability

The data supporting the results of this study are available upon request to the corresponding author.

Declarations

We have no competing interests to declare. We have no AI or LLM use to declare. The research proposal was reviewed and granted exemption by the University of Johannesburg Humanities Research Ethics Committee (REC-186-2022).

Authors' contributions

P.K.: Data analysis; writing – initial draft. K.D.: Writing – revisions, funding acquisition, project leadership. M.S.: Data curation; data analysis. D.M.: Data collection, sample analysis. All authors read and approved the final manuscript.

References

1. Dubé E, Vivion M, McDonald N. Vaccine hesitancy, vaccine refusal and the anti-vaccination movement: Influence, impact and implications. Expert Rev Vaccines. 2015;14(1):109–117. https://doi.org/10.1586/14760584.2015.964212
2. World Health Organization. Vaccine hesitancy: A growing challenge for immunization programs [webpage on the Internet]. c2015 [cited 2023 Nov 18]. Available from: www.who.int/news/item/18-08-2015-vaccine-hesitancy-a-growing-challenge-for-immunization-programmes
3. World Health Organization. Ten threats to global health in 2019 [webpage on the Internet]. c2019 [cited 2023 Nov 18]. Available from: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019
4. Cooper S, van Rooyen H, Wiysonge CS. COVID-19 vaccine hesitancy in South Africa: How can we maximize uptake of COVID-19 vaccines? Expert Rev Vaccines. 2021;20(8):921–933. https://doi.org/10.1080/14760584.2021.1949291
5. Dhanani LY, Franz B. A meta-analysis of COVID-19 vaccine attitudes and demographic characteristics in the United States. Public Health. 2022;207:31–38. https://doi.org/10.1016/j.puhe.2022.03.012
6. Africa CDC. COVID-19 vaccination [webpage on the Internet]. c2023 [cited 2024 Oct 17]. Available from: https://africacdc.org/covid-19/covid-19-vaccination/
7. Connoway K, Keal H, van Lill M, Joubert M. Increasingly polarised or finding common ground? Exploring pro- and anti-vaccine rhetoric on two South African Facebook pages. J Sci Comm. 2022;21(5):1–26. https://doi.org/10.22323/2.21050204
8. Gittings L, Casale M, Kannemeyer N, Rayalo N, Cluver L, Kelly J, et al. "Even if I'm well informed, I will never get it": COVID-19 vaccine beliefs, intentions and acceptability among adolescents and young people in South Africa. S Afr Health Rev. 2021;1:297–304. https://doi.org/10.61473/001c.75448
9. Hoffman B, Felter E, Chu KH, Shensa A, Hermann C, Wolynn T, et al. It's not all about autism: The emerging landscape of anti-vaccination sentiment on Facebook. Vaccine. 2019;37:2216–2223. https://doi.org/10.1016/j.vaccine.2019.03.003
10. Jokanovic I. A critical discourse analysis of the operation of power and ideology in COVID-19 vaccine Twitter commentary [Unpublished Master's thesis]. Johannesburg: University of the Witwatersrand; 2023.
11. Rozbroj T, Lyons A, Lucke J. Understanding how the Australian anti-vaccination movement perceives itself. Health Soc Care Commun. 2022;30(2):695–705. https://doi.org/10.1111/hsc.13182
12. Silubonde T, Knight L, Norris S, van Heerden A, Goldstein S, Draper C. Perceptions of the COVID-19 pandemic: A qualitative study with South African adults. BMC Public Health. 2023;23:684–698. https://doi.org/10.1186/s12889-023-15450-z
13. Watermeyer J, Scott M, Kapueja L, Ware LJ. To trust or not to trust: An exploratory qualitative study of personal and community perceptions of vaccines amongst a group of young community healthcare workers in Soweto, South Africa. Health Policy Plan. 2022;37:1167–1176. https://doi.org/10.1093/heapol/czac060
14. Hart PS, Chinn S, Soroka S. Politicization and polarization in COVID-19 news coverage. Sci Commun. 2020;42(5):679–697. https://doi.org/10.1177/1075547020950735
15. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101. https://doi.org/10.1191/1478088706qp063oa
16. Billig M, Condor S, Edwards D, Gane M, Middleton D, Radley A. Ideological dilemmas: A social psychology of everyday thinking. London: Sage; 1988.
17. Silverman D. Doing qualitative research. 6th ed. London: Sage; 2020.
18. Hoare J, Mendelsohn M, Frenkel L. COVID-19 vaccine hesitancy and anti-vaxxers – supporting healthcare workers to navigate the unvaccinated: Reflections from clinical practice. S Afr Med J. 2022;112(1):11–13. https://doi.org/10.7196/SAMJ.2022.v112i1.16208
19. Sunstein CR. Republic.com 2.0. Princeton, NJ: Princeton University Press; 2009.

AUTHORS:
Caradee Y. Wright[1,2]
Thandi Kapwata[1,3]
Nomfundo Mahlangeni[1,3]
Natasha Naidoo[1]
Candice Webster[1]

AFFILIATIONS:
[1]Climate Change and Health Research Programme, Environment and Health Research Unit, South African Medical Research Council, Pretoria, South Africa
[2]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa
[3]Department of Environmental Health, University of Johannesburg, Johannesburg, South Africa

CORRESPONDENCE TO:
Caradee Wright

EMAIL:
caradee.wright@mrc.ac.za

DATES:
Received: 20 Feb. 2024
Revised: 28 Jun. 2024
Accepted: 08 Aug. 2024
Published: 04 Dec. 2024

HOW TO CITE:
Wright CY, Kapwata T, Mahlangeni N, Naidoo N, Webster C. Assessing heat-related health perceptions in the minibus taxi industry in Tshwane, South Africa. S Afr J Sci. 2024;120(11/12), Art. #18030. https://doi.org/10.17159/sajs.2024/18030

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
Pfananani Ramulifho

KEYWORDS:
climate change, environmental health, global warming, health risks, heat stress

FUNDING:
South African National Research Foundation, South African Medical Research Council

Assessing heat-related health perceptions in the minibus taxi industry in Tshwane, South Africa

Global warming, the increase in heatwaves and periods of intense heat, is a major problem across the world, including in South Africa. Heat exposure has adverse health impacts, ranging from dehydration and heat stroke to death. The transport sector and its users are vulnerable to heat exposure both in vehicles and in places where they wait for transport. We investigated the heat perceptions and experiences of minibus commuters and the minibus taxi industry in minibus taxis and minibus taxi ranks to inform an intervention to prevent adverse health impacts from heat exposure in the City of Tshwane. Commuters (n = 279), drivers and marshals (n = 46) reported feeling hot in minibus taxis and minibus taxi ranks, experiencing sweating, headaches, exhaustion and dizziness. Some participants reported opening windows and doors, drinking water/cold drinks, removing a jersey and fanning themselves to try and cool down. All participants suggested solutions to address heat exposure in minibus taxis and minibus taxi ranks, including more shelters and trees for shade, the provision of water, benches and potentially an air-conditioned building for waiting/holding periods between trips. In consultation with our stakeholders, we prepared educational awareness materials about heat risks to health and actions to take when it is hot, and these were distributed through the minibus taxi sector in the City of Tshwane. Future research needs include measuring temperatures in these settings and rolling out large-scale interventions to protect health and well-being in a changing climate.

Significance:

Heat exposure has adverse health impacts, ranging from dehydration to death. The transport sector and its users are vulnerable to extreme heat. We investigated the heat experiences and perceptions of minibus commuters and the minibus taxi industry to inform an intervention to prevent adverse health impacts from heat exposure in the City of Tshwane. We prepared educational awareness materials about heat risks to health and actions to take when it is hot, and these were distributed through the minibus taxi sector. Future research needs include measuring temperatures in these settings and rolling out large-scale interventions to protect health and well-being in a changing climate.

Introduction

The Intergovernmental Panel on Climate Change (IPCC) Assessment Report 6 states that the average global temperature is projected to increase by 1.5–3 °C by 2050 and by 3–6 °C by 2100 compared to pre-industrial levels.[1] Extremely high temperatures and heatwaves have impacts on human health, including heat stroke, dehydration and respiratory problems, among other adverse effects.[2] Given these projections, it is crucial to study heat-related health impacts across various daily scenarios.

Heat exposure can be problematic in specific settings, such as in schools[3,4], outdoor work environments[5] and primary healthcare facilities[6]. It is also increasingly being recognised as a problem in the transport sector, especially inside public transport vehicles and at terminals/ranks.[7] Approximately 80% of the urban population in Africa use some form of public transport (minibus taxis, buses, moto-taxis, matatus, sept-place taxis, trotros, keke, danfo, boda boda, etc.).[8,9] There is a scarcity of studies that explore the health repercussions of heat exposure on populations using different modes of transport, particularly in Africa. Studies exploring the effects of climate change on health in the transport sector have focused on urban and transport planning predominantly in high-income countries[10] and on the analysis of air-pollutant emissions[11-14]. One area of research on heat-related impacts within the sector includes mortality of infants and young children left unattended in cars during periods of intense heat or on hot days in high-income countries[15-19] and in low- to middle-income countries[17,20]. However, there is a lack of studies examining reported deaths of this nature in Africa. A study analysing disease burden in the urban areas of Barcelona in Spain found that reducing motor vehicle usage, implementing green infrastructure and encouraging physical activity could enhance public health outcomes and reduce air pollution and heat exposure in cities, highlighting the potential benefits of similar interventions in the minibus taxi industry to mitigate heat exposure and improve health in urban settings.[21] In Phoenix, Arizona, in the USA, bus-stop infrastructure failed to adequately shield commuters from extreme heat[7], and planting trees near bus-stops were shown to lessen health risks[22]. There is an urgent need for comprehensive studies on the impact of heat exposure in the transport sector to inform effective policies and interventions targeting adverse health effects.

Public transportation contributes significantly to global efforts to address climate change by reducing the number of vehicles on the road.[23] In addition, it is more affordable for families with lower incomes to use public transport. Thus, it remains imperative to improve public transport service delivery to meet the needs of the poorest populations.[24] A recent study spanning Uganda, Zambia and 15 other African countries highlighted the vital role of climate resilience and inclusive access to transport, emphasising the need for strong political will, cross-sector collaboration, inclusivity for vulnerable groups, and sufficient funding to develop climate-resilient

transport infrastructure across Africa.[25] Further studies are imperative to understand the health outcomes of heat exposure in transport systems, especially in Africa, to integrate climate resilience and inclusive transport infrastructure that meets the needs of vulnerable populations.

Despite the transport sector's heavy reliance on fossil fuels (99%), South Africa does not prioritise national adaptation strategies for this sector, despite the urgent need for decarbonisation in both passenger and freight transport.[26] In South Africa, 67% of public transport passengers use minibus taxis, and the industry is the most popular means of transport in South Africa, surpassing both trains (13%) and buses (20%).[27] The majority of users are the low-income population, who live in townships that are not well served by formal transportation, such as buses and trains.[28] Across the country, the minibus taxi industry has around 150 000 minibus taxi owners operating between 200 000 and 250 000 minibus taxis.[8] In addition, the minibus taxi industry employs about 300 000 drivers, 100 000 rank marshals, 100 000 car washers and 150 000 street sellers at minibus taxi ranks.[8]

Given the number of role players in the minibus taxi sector potentially exposed to extreme heat, this study aimed to understand the heat experiences and perceptions of commuters and employees of the minibus taxi industry to inform an intervention to mitigate against the adverse health impacts of heat exposure. Our objectives were to (1) conduct a survey among minibus taxi commuters and drivers to understand their experiences of heat exposure and (2) explore interventions they suggest could be implemented to help prevent adverse heat-related health impacts in minibus taxis and at taxi ranks. To the best of our knowledge, this is the first study to assess heat-related health perceptions among minibus taxi commuters, marshals and drivers in South Africa, and possibly Africa.

Methods

Study site

The site for the administration of the questionnaire survey was purposefully chosen, primarily for safety reasons. In collaboration with the City of Tshwane (CoT) and the Tshwane Taxi Industry (TTI), the Castle Gate Taxi rank (Figure 1) was deemed an appropriate, safe location to conduct the survey. It was acknowledged that this taxi rank was different from other ranks in that it had shade structures and seating, unlike some other ranks in the City; however, the commuters and drivers would be similar to those visiting other ranks.

Procedures

We carried out a cross-sectional survey study to assess heat-related experiences among commuters and the taxi industry. Fieldworkers spent 5 days (5–9 December 2022) at the Castle Gate minibus taxi rank and recruited people to participate in the survey. Participants were provided with an information sheet and provided informed consent prior to participation in the survey. Participants needed to be 18 years or older to participate in the study, and no other exclusion criteria were applied.

The questionnaire consisted of several questions:

- What gender are you (male, female)?
- What age are you (18–25 years, 26–35 years, 36–45 years, 46–55 years, 56–65 years, older than 66 years)?
- When travelling in a minibus taxi, do you feel hot (no, yes)?
- Do you feel hot when travelling in a minibus taxi at 11:00 in the morning (no, yes)?
- Do you feel hot when travelling in a minibus taxi at 15:00 in the afternoon (no, yes)?
- When you feel hot, do you do something to try and feel cooler (no, yes)?
- What do you do to try and feel cooler when you feel hot in a minibus taxi (open answers)?
- Have you ever experienced any of the following symptoms when you feel hot while in the taxi in the morning (sweating, heat cramps, headaches, a faster heartbeat, nausea/vomiting, heat rash / pimples or blisters on the skin, heat exhaustion / feeling tired or weak, dizziness / confusion, difficulties breathing, fainting and / or irritability[29])?
- Have you ever experienced any of the following symptoms when you feel hot while in the taxi in the afternoon (sweating, heat cramps, headaches, a faster heartbeat, nausea/vomiting, heat rash / pimples or blisters on the skin, heat exhaustion / feeling tired or

Figure 1: Minibus taxis parked at the Castle Gate Shopping Centre in Tshwane, South Africa.

weak, dizziness / confusion, difficulties breathing, fainting and / or irritability[29])?

- In hot weather, when you are standing at or walking through the minibus taxi rank, do you feel hot (no, yes)?
- Have you ever experienced any of the following symptoms when you feel hot while in the minibus taxi rank (sweating, heat cramps, headaches, a faster heartbeat, nausea/vomiting, heat rash / pimples or blisters on the skin, heat exhaustion / feeling tired or weak, dizziness / confusion, difficulties breathing, fainting and / or irritability[29])?
- What do you think can be done to make minibus taxis and minibus taxi ranks cooler and more comfortable when it's hot (open answers)?

Statistics

Fieldworkers used tablets to record answers from study participants. Participants' answers were recorded using REDCap[30,31] software installed on each tablet. REDCap is a secure web application designed for the collection and management of research data.[31] It is a versatile and efficient tool, widely adopted by researchers to streamline their data workflows and ensure high-quality, compliant research practices. All data were processed for quality control prior to being imported into Stata version 10[7] for analysis. Descriptive statistics were calculated, and open answers were collated by themes identified from the participants' responses. The results of this study will serve as a basis for future studies examining the heat-health effects when travelling or working in minibus taxis. Thus, descriptive statistics were used for the effectiveness in summarising and presenting the survey data, ensuring data quality, and providing actionable insights into the heat-related experiences of commuters and the taxi industry.

Results

A total of 324 participants (279 commuters and 46 members of the minibus taxi industry) completed the questionnaire survey (Table 1). Most participants reported that they felt hot when travelling in a minibus taxi, regardless of whether it was in the morning or the afternoon. Participants said that when they felt hot inside a minibus taxi, they would open a window (87%), drink water (78%) or cold drinks (15%), fan themselves with their hand or a make-shift fan, and remove a jersey.

Results showed that, among the 324 participants, the most common health complaints when it was hot were sweating (~between 50% and 80%), heat exhaustion / feeling tired or weak (~20%), headaches (~30%) and dizziness (20%).

Heat exposure and health effects among commuters

A substantial percentage of commuters (75%) reported feeling hot sometimes when travelling in a minibus taxi, with an additional 22% always feeling hot. At 11:00 in the morning, 65% of commuters sometimes felt hot, while at 15:00 in the afternoon, the percentage of those who sometimes or always felt hot increased to 94%. To cool down, 81% of commuters reported taking some action, with 87% opening a window and 78% drinking water. Common heat illness symptoms experienced in the morning included sweating (72%), headaches (27%) and dizziness/confusion (16%) (Figure 2). In the afternoon, sweating (77%), headaches (32%) and dizziness/confusion (19%) remained prevalent (Figure 2). When standing at or walking through the minibus taxi rank in hot weather, 89% felt hot, experiencing symptoms such as sweating (84%), headaches (32%) and heat exhaustion (21%) (Figure 2).

Heat exposure and health effects among drivers

When travelling in a minibus taxi, 63% of drivers reported sometimes feeling hot, and 30% always felt hot. At 11:00 in the morning, 65% sometimes felt hot, whereas, at 15:00 in the afternoon, this reduced slightly to 50%. To alleviate the heat, 93% of drivers took actions such as drinking water (89%) and opening a window (82%). In terms of heat illness symptoms in the morning, sweating (56%) and dizziness/confusion (19%) were most common. While in the taxi ranks, 93% of drivers reported feeling hot, with sweating (67%), headaches (17%) and heat exhaustion (21%) being the predominant symptoms. These findings indicate that heat exposure is a significant concern for drivers, affecting their comfort and health both during transit and while waiting at the ranks.

Unfortunately, for the question 'Have you ever experienced any of the following symptoms when you feel hot while in the minibus taxi in the afternoon?' posed to the taxi industry drivers, the data were not recorded and are missing.

Comparative analysis between the experiences of drivers and commuters

There are notable differences and some similarities between the experiences of drivers and commuters regarding heat exposure and its effects. All 46 drivers surveyed were men, whereas commuters were more gender diverse (54% male and 46% female). Drivers were more evenly distributed across age groups, while most commuters were aged 26–35 years (53%). A larger percentage of drivers (30%) reported 'always feeling hot while travelling in a minibus taxi' compared to commuters (22%). At 11:00 in the morning, 20% of drivers 'never felt hot', compared to only 12% of commuters. In the afternoon at 15:00, 41% of drivers

Table 1: Results of the questionnaire administered to the minibus taxi commuters, marshals and the drivers ($N = 324$)

	Commuters $N = 279$		Drivers $N = 46$	
	Frequency		Frequency	
Question	**Number (*n*)**	**Percentage (%)**	**Number (*n*)**	**Percentage (%)**
Gender				
Male	152	54	46	100
Female	127	46	0	0
Age				
18–25 years	41	15	1	2
26–35 years	148	53	13	28
36–45 years	66	23	14	30
46–55 years	17	6	13	28
56–65 years	6	2	4	8
Older than 66 years	1	<1	1	2

...Table 1 continues on next page

Table 1 continued...

	Commuters $N = 279$		Drivers $N = 46$	
	Frequency		Frequency	
Question	**Number (*n*)**	**Percentage (%)**	**Number (*n*)**	**Percentage (%)**
When travelling in a minibus taxi, do you feel hot?				
Never	9	3	3	7
Sometimes	208	75	29	63
Always	61	22	14	30
Missing	1	<1	0	0
Do you feel hot when travelling in a minibus taxi at 11:00 in the morning?				
Never	34	12	9	20
Sometimes	176	65	28	65
Always	60	22	6	13
Missing	9	3	3	6
Do you feel hot when travelling in a minibus taxi at 15:00 in the afternoon?				
Never	16	6	1	2
Sometimes	101	38	23	50
Always	149	56	19	41
Missing	13	4	3	6
When you feel hot, do you do something to try and feel cooler?				
No	6	2	3	7
Yes	228	81	43	93
Missing	45	45	0	0
What do you do to cool down when it's hot in a minibus taxi?				
Drink water	218	78	41	89
Drink cold drinks / fizzy drinks	44	15	10	21
Open a window	244	87	38	82
Other (discussed in text)				
Have you ever experienced any of the following symptoms when you feel hot while in a minibus taxi in the morning?				
Sweating	200	72	26	56
Heat cramps	11	4	3	6
Headaches	76	27	6	13
A faster heartbeat	12	4	0	0
Nausea/vomiting	24	8	1	2
Heat rash / pimples or blisters on the skin	21	8	5	10
Heat exhaustion / feeling tired or weak	46	16	5	10
Dizziness / confusion	45	16	9	19
Difficulties breathing	21	8	2	4
Fainting and / or irritability	6	2	0	0

...Table 1 continues on next page

Table 1 continued...

	Commuters $N = 279$		Drivers $N = 46$	
	Frequency		Frequency	
Question	**Number (*n*)**	**Percentage (%)**	**Number (*n*)**	**Percentage (%)**
Have you ever experienced any of the following symptoms when you feel hot while in the minibus taxi in the afternoon?				
Sweating	215	77	–#	–#
Heat cramps	11	4	–	–
Headaches	90	32	–	–
A faster heartbeat	20	7	–	–
Nausea/vomiting	24	8	–	–
Heat rash / pimples or blisters on the skin	26	9	–	–
Heat exhaustion / feeling tired or weak	62	22	–	–
Dizziness / confusion	54	19	–	–
Difficulties breathing	16	6	–	–
Fainting and / or irritability	6	2	–	–
In hot weather, when you are standing at or walking through the minibus taxi rank, do you feel hot?				
No	31	11	3	7
Yes	244	89	43	93
Have you ever experienced any of the following symptoms when you feel hot while in the minibus taxi rank?				
Sweating	236	84	31	67
Heat cramps	16	5	2	4
Headaches	91	32	8	17
A faster heartbeat	15	5	1	2
Nausea/vomiting	16	6	3	6
Heat rash / pimples or blisters on the skin	31	11	4	8
Heat exhaustion / feeling tired or weak	58	21	10	21
Dizziness / confusion	54	19	7	15
Difficulties breathing	20	7	2	4
Fainting and / or irritability	4	1	3	3

Note: # missing data

'always felt hot', compared to 56% of commuters. A high percentage of both drivers (93%) and commuters (81%) reported taking action to cool down. Common actions included drinking water (89% drivers, 78% commuters) and opening windows (82% drivers, 87% commuters). Both groups reported sweating as the most common symptom in the morning (56% drivers, 72% commuters), followed by headaches (13% drivers, 27% commuters) and dizziness/confusion (19% drivers, 16% commuters). Only commuters' data are available for afternoon symptoms, showing high incidences of sweating (77%) and headaches (32%). A larger proportion of drivers (93%) felt hot at taxi ranks compared to commuters (89%). Symptoms at ranks were similar, with sweating being the most common (67% drivers, 84% commuters), followed by headaches (17% drivers, 32% commuters) and heat exhaustion (21% for both groups).

Participant suggestions for improving comfort in minibus taxis and taxi ranks

Several responses were provided to the question 'What do you think can be done to make minibus taxis and minibus taxi ranks cooler and more comfortable when it's hot?'. These included: adding shade / more shelters with benches, extending existing shelters, planting more trees, providing access to drinking water/taps, providing a building with air conditioning to wait inside, and having water mist sprays.

Participants also suggested eating ice, keeping the minibus taxi door open while the minibus taxi is waiting at the minibus taxi rank, using an umbrella for shade when waiting for a minibus taxi, and spending time looking at WhatsApp messages.[4,32]

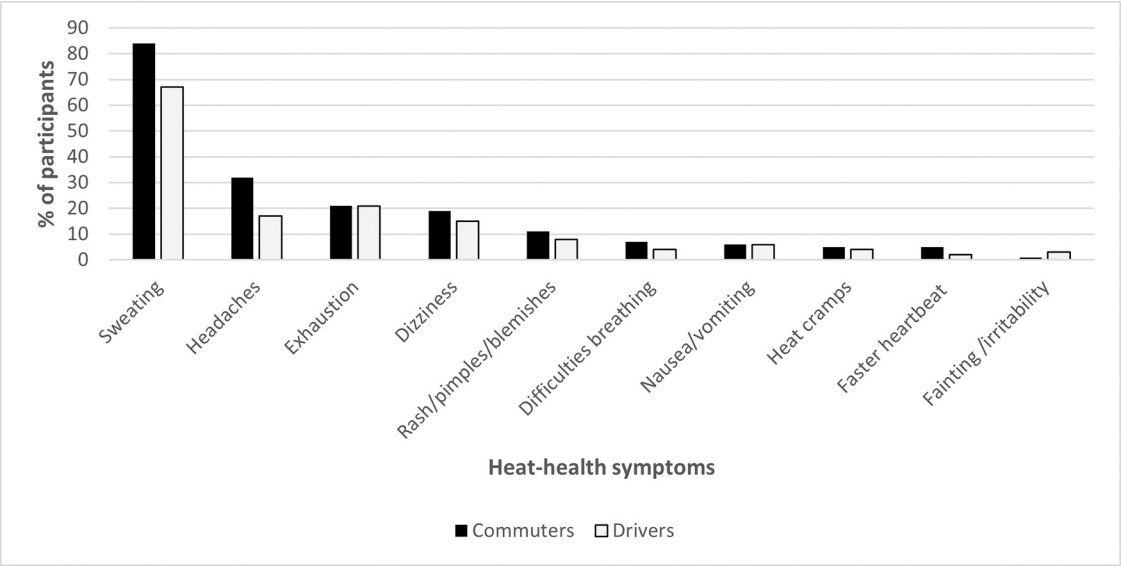


Figure 2: Reported heat-health symptoms among participants in the study.

Discussion

Despite the improvement of life expectancy in Africa, the continent still has the highest rates of many preventable, non-communicable, infectious diseases and chronic illnesses.[33] Low-income populations are widely recognised as being particularly vulnerable to adverse health effects arising from extreme heat.[34] Studying the effects on all African communities that could potentially be exposed to heat-related health problems is therefore imperative. We set out to understand heat-related health perceptions among minibus taxi commuters, marshals and drivers in the CoT.

We found that heat exposure is perceived as a problem that has impacts on human health. Moreover, people have found ways to adapt to heat exposure when in minibus taxis and ranks, although they still experience adverse heat-related symptoms, suggesting that more should be done to protect them from heat-related impacts. People in this transport sector also have many sector-relevant suggestions on how to improve heat-related conditions in minibus taxis and ranks. Moreover, despite the existing shading at the study site, commuters indicated a need for additional trees and shade structures to address the persistent heat-related issues. Thus, enhancing shading infrastructure across all ranks is essential to mitigate heat exposure effectively, especially at taxi ranks that lack trees and shading.

A crucial part of our analysis was the perceptions of taxi drivers in the study, as these individuals are exposed to extreme heat in taxis throughout the day. A previous study analysing the health of taxi drivers in Tshwane showed that they are prone to obesity, hypertension and type II diabetes, linked to unhealthy lifestyle habits.[35,36] Importantly, heat illness has been shown to exacerbate underlying conditions and chronic illnesses such as respiratory and cardiovascular disease.[37,38] Our results on the taxi drivers heat-health perceptions underscore the significant risk imposed on taxi drivers' overall health and well-being. In addition, the health and thermal comfort of taxi drivers ultimately affects commuters who sit in the taxis being driven by them.

Previous studies elsewhere in the world have considered different types of commuters and environmental health risks.[7,22,39] For instance, in the USA, the number of commuters on subways was lower under extreme temperatures compared to cooler days.[40] In South Africa, where so many households rely on minibuses for transport, it is unlikely that their use will decline. Instead, we need to 'heat-proof' the industry, including inside minibus taxis and ranks. We also need to raise awareness among communities about the risks of heat exposure and actions that can be taken to reduce risks.

While our study provides valuable insights into the prevalence and nature of heat-related symptoms among minibus taxi passengers, it is important to acknowledge certain limitations. One key limitation is the potential for misattribution of symptoms to heat exposure. Although the questionnaire was designed to ask participants about symptoms experienced 'when they feel hot in the minibus taxi', it is possible that some reported symptoms could have other underlying causes not related to heat exposure. The reliance on self-reported data may introduce bias, as participants might not accurately recall or attribute their symptoms correctly. Furthermore, while the symptoms listed in our questionnaire are recognised indicators of heat illness (sweating, heat cramps, headaches, faster heartbeat, nausea/vomiting, heat rash, heat exhaustion, dizziness/confusion, difficulties breathing, fainting and irritability), they can also be associated with other conditions. Therefore, care should be taken not to draw definitive causal relationships between the reported symptoms and heat exposure based solely on this questionnaire. To address this limitation, future studies could incorporate objective measures of heat exposure and physiological responses, such as temperature and heart rate monitoring, to more accurately determine the impact of heat on passengers' health. Additionally, longitudinal studies could help establish clearer causal links between heat exposure and specific health outcomes.

Participants suggested several interventions to make minibus taxis and taxi ranks cooler, including more shelters with benches, planting trees, installing water taps, providing air-conditioned waiting areas, and using water mist sprays. Notably, some participants recommended using WhatsApp for disseminating information, which sparked the idea for an education campaign and awareness materials (Figure 3). We proposed all the participants' suggestions for interventions to the CoT and the TTI, each with its pros and cons (Table 2). The CoT and TTI recommended an education campaign to raise awareness about heat-related health risks and mitigation strategies for minibus taxi commuters and drivers. This led to the development of the 'Throw Shade on Heat this Summer' campaign, which included five materials: a car licence sticker, a pull-out banner, a long sticker for above the minibus taxi door (Figure 3), a 2024 calendar, and a four-page flyer. The content was derived from the existing literature[4,32] and previous studies in South Africa and emphasised carrying water, wearing loose clothing, opening windows while travelling, recognising heat illness symptoms, using umbrellas or hats, and caring for vulnerable individuals. These materials were distributed to minibus taxi commuters, drivers, marshals and owners by the CoT and TTI. Although the materials were well received, planned follow-up focus group discussions with the CoT and TTI were not possible due to unforeseen circumstances affecting the relationship between the parties at that time.

Figure 3: The long sticker designed to go above the main, large door of a minibus taxi.

Table 2: Pros and cons of each of the participant's suggested interventions to combat heat in minibus taxis and minibus taxi ranks

Intervention	Pros	Cons
More shelters / extend shelters	Reliable shade	Costly
	Also rainproof	May not cast shade in the appropriate place
	Accommodates many people	May be vandalised
	Could paint roofs white to reflect heat	May attract homeless people
More trees	Casts shade all year round (evergreen trees)	Roots damage tar and paving
	Well-established, mature trees are available	Need water and fertiliser
		Might be cut down for firewood
Access to drinking water / ice	Best way for people to stay healthy during hot weather is to drink water	Costly
	Accessible water encourages people to drink more water	Difficult to maintain infrastructure
	Drinking fountains exist	Difficult to ensure clean water and a continuous supply
		People may take a look to meet daily needs rather than just to quench thirst
		Difficult to protect
		People need a bottle or cup
Keep minibus taxi door open when minibus taxi is waiting at the rank / holding zone	Simple	Safety and security concerns
	Easy	Exposure to air pollution
	Allows through-breeze	Driver's choice
		Might not help if the weather is very hot
		Difficult to enforce
Receive a heat-related health warning	Messages are easy to make and disseminate	Regularity can lead to messages being 'normalised' and ignored
	Can be in multiple languages	Some actions may be difficult, e.g. water accessibility is an issue
	Prompts people to act, e.g. open a window, remove a jersey	

Future research is required to measure temperatures in minibus taxis and minibus taxi ranks to quantitatively assess heat-related health risks to commuters, marshals and drivers. Moreover, more elaborate interventions, such as infrastructural changes, are likely needed to protect the health and well-being of the residents of the CoT, and communities elsewhere, from the risks associated with heat exposure. In addition, intervention studies have the potential to inform urban planning, design and climate change adaptation strategies.

Acknowledgements

We thank the City of Tshwane, the Tshwane Taxi Industry, Castle Gate Mall and all our participants.

Funding

This project was funded by the National Research Foundation and the South African Medical Research Council.

Data availability

The data supporting the results of this study are available upon request to the corresponding author.

Declarations

We have no competing interests to declare. We declare that no AI was used in the production of this manuscript. Research ethics clearance for the study was granted by the South African Medical Research Council Ethics Committee (EC019-9/2022). Each participant was given a ZAR150 shopping voucher for their participation, and hence, we were holding multiple vouchers on each day of the study. A requirement of the ethics approval was that participants be provided with a token of appreciation in the form of a shopping voucher.

Authors' contributions

C.Y.W.: Conceptualisation, methodology, data collection, sample analysis, validation, data curation, writing – the initial draft, writing revisions, student supervision, project leadership, project management, funding acquisition. T.K.: Methodology, student supervision. N.M.: Student supervision. N.N.: Student supervision. C.W.: Student supervision. All authors read and approved the final manuscript.

References

1. Trisos CH, Adelekan IO, Totin E, Ayanlade A, Efitre J, Gemeda A, et al. Chapter 9: Africa. In: Pörtner HO, Roberts DC, Tignor M, Poloczanska ES, Mintenbeck K, Alegría A, et al., editors. Climate change 2022: Impacts, adaptation and vulnerability. Contribution of Working Group II to the sixth assessment report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press; 2023. p. 1285–1456. https://doi.org/10.1017/9781009325844.011
2. Arsad FS, Hod R, Ahmad N, Ismail R, Mohamed N, Baharom M, et al. The impact of heatwaves on mortality and morbidity and the associated vulnerability factors: A systematic review. Int J Environ Res Public Health. 2022;19(23), Art. #16356. https://doi.org/10.3390/ijerph192316356
3. Shortridge A, Walker W VI, White DD, Guardaro MM, Hondula DM, Vanos JK. HeatReady schools: A novel approach to enhance adaptive capacity to heat through school community experiences, risks, and perceptions. Clim Risk Manag. 2022;36, Art. #100437. https://doi.org/10.1016/j.crm.2022.100437
4. Bidassey-Manilal S, Wright C, Engelbrecht J, Albers P, Garland R, Matooane M. Students' perceived heat-health symptoms increased with warmer classroom temperatures. Int J Environ Res Public Health. 2016;13(6), Art. #566. https://doi.org/10.3390/ijerph13060566
5. Moda HM, Filho WL, Minhas A. Impacts of climate change on outdoor workers and their safety: Some research priorities. Int J Environ Res Public Health. 2019;16(18), Art. #3458. https://doi.org/10.3390/ijerph16183458
6. Lokotola CL, Mash R, Naidoo K, Mubangizi V, Mofolo N, Schwerdtle PN. Climate change and primary health care in Africa: A scoping review. J Clim Change Health. 2023;11, Art. #100229. https://doi.org/10.1016/j.joclim.2023.100229
7. Dzyuban Y, Hondula DM, Coseo PJ, Redman CL. Public transit infrastructure and heat perceptions in hot and dry climates. Int J Biometeorol. 2022;66(2):345–356. https://doi.org/10.1007/s00484-021-02074-4
8. Competition Commission of South Africa (CCSA). Market inquiry into the land based public passenger transport sector- main report non-confidential [webpage on the Internet]. c2021 [cited 2024 Mar 16]. Available from: https://www.compcom.co.za/public-passenger-transport-market-inquiry-2/
9. Agyei-Boakye OP. Public transportation in Africa: Informal transport or paratransit [webpage on the Internet]. c2022 [cited 2024 Jun 25]. Available from: https://www.transportafrica.org/public-transportation-in-africa-informal-transport-or-paratransit
10. Nieuwenhuijsen MJ. Urban and transport planning pathways to carbon neutral, liveable and healthy cities: A review of the current evidence. Environ Int. 2020;140, Art. #105661. https://doi.org/10.1016/j.envint.2020.105661
11. Jones S, Tefe M, Zephaniah S, Tedla E, Appiah-Opoku S, Walsh J. Public transport and health outcomes in rural sub-Saharan Africa – A synthesis of professional opinion. J Transp Health. 2016;3(2):211–219. https://doi.org/10.1016/j.jth.2015.12.005
12. Luo Z, Wang Y, Lv Z, He T, Zhao J, Wang Y, et al. Impacts of vehicle emission on air quality and human health in China. Sci Total Environ. 2022;813, Art. #152655. https://doi.org/10.1016/j.scitotenv.2021.152655
13. Ekpenyong CE, Ettebong EO, Akpan EE, Samson TK, Daniel NE. Urban city transportation mode and respiratory health effect of air pollution: A cross-sectional study among transit and non-transit workers in Nigeria. BMJ Open. 2012;2(5), e001253. https://doi.org/10.1136/bmjopen-2012-001253
14. Anenberg S, Miller J, Henze D, Minjares R. A global snapshot of the air pollution-related health impacts of transportation sector emissions in 2010 and 2015, International Council on Clean Transportation (ICCT) [webpage on the Internet]. c2021 [cited 2024 Mar 16]. Available from: https://theicct.org/publication/a-global-snapshot-of-the-air-pollution-related-health-impacts-of-transportation-sector-emissions-in-2010-and-2015/
15. Ho K, Minhas R, Young E, Sgro M, Huber JF. Paediatric hyperthermia-related deaths while entrapped and unattended inside vehicles: The Canadian experience and anticipatory guidance for prevention. Paediatr Child Health. 2020;25(3):143–148. https://doi.org/10.1093/pch/pxz087
16. Hammett DL, Kennedy TM, Selbst SM, Rollins A, Fennell JE. Pediatric heatstroke fatalities caused by being left in motor vehicles. Pediatr Emerg Care. 2021;37(12):e1560–e1565. https://doi.org/10.1097/PEC.0000000000002115
17. Costa D, Grundstein A. An analysis of children left unattended in parked motor vehicles in Brazil. Int J Environ Res Public Health. 2016;13(7), Art. #649. https://doi.org/10.3390/ijerph13070649
18. National Safety Council. Hot car deaths [webpage on the Internet]. c2023 [cited 2024 Mar 16]. Available from: https://injuryfacts.nsc.org/motor-vehicle/motor-vehicle-safety-issues/hotcars/
19. Dowd MD. Vehicular hyperthermia – A highly preventable and potentially fatal problem. Pediatr Ann. 2018;47(3):e88–e90. https://doi.org/10.3928/19382359-20180220-04
20. Siddiqui SA, Siddiqui GF, Singh MV. Vehicular hyperthermia deaths in Indian children: A little recognised mode of fatal injury. Trop Doct. 2021;51(1):109–111. https://doi.org/10.1177/0049475520945445
21. Mueller N, Rojas-Rueda D, Basagaña X, Cirach M, Cole-Hunter T, Dadvand P, et al. Health impacts related to urban and transport planning: A burden of disease assessment. Environ Int. 2017;107:243–257. https://doi.org/10.1016/j.envint.2017.07.020
22. Lanza K, Durand CP. Heat-moderating effects of bus stop shelters and tree shade on public transport ridership. Int J Environ Res Public Health. 2021;18(2), Art. #463. https://doi.org/10.3390/ijerph18020463
23. Welle B, Kustar A, Tun TH, Albuquerque C. Post-pandemic, public transport needs to get back on track to meet global climate goals [webpage on the Internet]. c2023 [cited 2024 Jun 25]. Available from: https://www.wri.org/insights/current-state-of-public-transport-climate-goals
24. Kaiser N, Barstow CK. Rural transportation infrastructure in low- and middle-income countries: A review of impacts, implications, and interventions. Sustainability. 2022;14(4), Art. #2149. https://doi.org/10.3390/su14042149
25. Cinderby S, Haq G, Opiyo R, Muhoza C, Ngabirano A, Wasike Y, et al. Inclusive climate resilient transport challenges in Africa. Cities. 2024;146, Art. #104740. https://doi.org/10.1016/j.cities.2023.104740
26. World Bank Group. Comparing G20 climate action towards net zero: Climate transparency report [webpage on the Internet]. c2021 [cited 2024 Jun 25]. Available from: www.climate-transparency.org
27. Behrens R. Bottlenecks, delays and a few street-smart solutions. In: Intelligence Report: No public transport shortcuts for SA [webpage on the Internet]. c2020 [cited 2024 Mar 16]. Available from: https://www.oldmutual.co.za/v3/assets/blt0554f48052bb4620/bltee8c6c07e4b2085d/5f69d205a600d154600e5f14/transport-intelligence-report_july-2020_final.pdf
28. Vegter I. South Africa's minibus taxi industry: Resistance to formalisation and innovation. Johannesburg: South African Institute of Race Relations; 2020.
29. Leiva DF, Church B. Heat Illness, In: StatPearls [webpage on the Internet]. c2023 [cited 2024 Mar 16]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK553117/
30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377–381. https://doi.org/10.1016/j.jbi.2008.08.010

31. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019;95, Art. #103208. https://doi.org/10.1016/j.jbi.2019.103208

32. Wright C, Street R, Cele N, Kunene Z, Balakrishna Y, Albers P, et al. Indoor temperatures in patient waiting rooms in eight rural primary health care centers in northern south africa and the related potential risks to human health and wellbeing. Int J Environ Res Public Health. 2017;14(1), Art. #43. https://doi.org/10.3390/ijerph14010043

33. Health in Africa. Nat Commun. 2024;15(1), Art. #967. https://doi.org/10.1038/s41467-024-45268-1

34. 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

35. Ramukumba TS, Mathikhi MS. Health assessment of taxi drivers in the city of Tshwane. Curationis. 2016;39(1), Art. #a1671. https://doi.org/10.4102/curationis.v39i1.1671

36. Mabetwa EM, Mokwena KE, Mphekgwana PM, Modjadji P. Metabolic syndrome and its components among taxi drivers in the city of Tshwane, South Africa. Appl Sci. 2022;12(3), Art. #1767. https://doi.org/10.3390/app12031767

37. 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

38. Alahmad B, Khraishah H, Shakarchi AF, Albaghdadi M, Rajagopalan S, Koutrakis P, et al. Cardiovascular mortality and exposure to heat in an inherently hot region: Implications for climate change. Circulation. 2020;141(15):1271–1273. https://doi.org/10.1161/CIRCULATIONAHA.119.044860

39. Lin TP, Hwang RL, Huang KT, Sun CY, Huang YC. Passenger thermal perceptions, thermal comfort requirements, and adaptations in short- and long-haul vehicles. Int J Biometeorol. 2010;54(3):221–230. https://doi.org/10.1007/s00484-009-0273-9

40. Stechemesser A, Wenz L. Inequality in behavioural heat adaptation: An empirical study with mobility data from the transport system in New York City, NY, USA. Lancet Planet Health. 2023;7(10):e798–e808. https://doi.org/10.1016/S2542-5196(23)00195-X

Mobile phone based laser scanning as a low-cost alternative for multidisciplinary data collection

AUTHORS:
Kian J. Jeftha[1]
Moreblessings Shoko[1]

AFFILIATION:
[1]Geomatics Division, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Moreblessings Shoko

EMAIL:
m.shoko@uct.ac.za

DATES:
Received: 12 Jan. 2023
Revised: 24 Mar. 2024
Accepted: 04 Sep. 2024
Published: 04 Dec. 2024

HOW TO CITE:
Jeftha KJ, Shoko M. Mobile phone-based laser scanning as a low-cost alternative for multidisciplinary data collection. S Afr J Sci. 2024;120(11/12), Art. #15437. https://doi.org/10.17159/sajs.2024/15437

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITORS:
Michael Inggs
Thywill Dzogbewu

KEYWORDS:
multidisciplinary, mobile technologies, laser scanning, low cost, iLiDAR, sustainable solutions

FUNDING:
None

Airborne and terrestrial laser scanners have traditionally been used as specialised toolsets for three-dimensional scene capture in engineering, providing highly accurate measurements with increasingly minimal human interaction. However, commercial or engineering-grade scanning instruments remain expensive and sensitive, requiring costly routine calibrations to ensure their optimum functionality. The recent inclusion of laser scanning sensors by mobile phone corporations such as Apple Computer Inc. is now analogous to the integration of Global Navigation Satellite Systems (GNSS) and cameras into smartphones as seen decades ago. Likely, these initial efforts to include the scanning sensor in mobile phones will see rapid improvements in the application and accuracy of the sensor to serve the growing need for scanning applications for transdisciplinary users. However, there is a limited amount of literature that benchmarks the emerging and low-cost scanning sensors to existing commercial ones to inform practice, thus prompting a need for researchers to evaluate and provide scientific evidence that can inform multidisciplinary scanning. It was noted that there was some absolute positional shift and scan drift in the iPhone Light Detection and Ranging (LiDAR) data. The researchers therefore investigated the extent to which the accuracy of laser scanning tools available within the iPhone 12 Pro compared to engineering-grade laser scanners. Outcomes from the study showed that iPhone scanners can deliver the required models, despite being unstable in dynamic environments when pitched against engineering-grade LiDAR scanners. The research recommends that stabilisers, such as stabilising gimbals or enhanced GNSS receivers, be used in practice to achieve improved accuracy from the mobile phone LiDAR.

Significance:

Laser scanners offer multiple advantages for modelling features in three dimensions in diverse applications, including documentation, archaeology, environmental modelling and mapping. However, the cost of entry to acquire scan data has been a limitation to its wide-scale use across multiple disciplines. This study demonstrated from an accuracy-based perspective that iPhone scanners can deliver the required models addressing different model purposes, despite being slightly unstable when pitched against engineering-grade Light Detection and Ranging (LiDAR) scanners. The results are significant in reinforcing the competence of low-cost tools in increasing access and use of this technology in curating three-dimensional models for multidisciplinary work.

Introduction, background and aims

Light Detection and Ranging (LiDAR) scanning systems allow users to create observations of any human-constructed or environmental structures for application in hundreds of areas such as geology, real estate, heritage, archaeology, deformation monitoring, engineering and other precise spatial data collection.[1,2] Scanning approaches a broad spectrum of users across the technical context of aerial (ALS), mobile (MLS) and terrestrial laser scanning (TLS) applications.[1-3] These scanning applications (which can be airborne, mobile or terrestrial) are known to provide quick and accurate multi-point positional data, which are currently collected at increasingly faster rates.[1,3] Furthermore, terrestrial laser scanning, in particular, employs the use of high-grade equipment capable of providing excessively large volumes of point cloud data with detailing for topographic mapping, meteorology, archaeology, deformation monitoring, construction and mechanical structure analysis in engineering.[1,3] This diverse range of applications has also been met by rapid hardware developments and processing capabilities, with technology developers attempting to make their offerings more accurate, more efficient and more affordable to a larger market.[4,5] Recent applications have also extended to include human and social sciences as well as arts where physical models are now increasingly of interest.[5-7] Typical examples include metrology, health sciences, development studies, forensics, biodiversity and many others. However, despite increased awareness and the need for scan technology, actual access to scanning services has remained a barrier to its full-scale or broad application due to the prohibitive costs associated with commercial scan equipment and its maintenance plans.[8,9]

On the contrary, since 2007, when the first iPhone was announced, the world has seen a rise in related high-end mobile phone interfaces that have claimed to revolutionise many professional disciplines, by offering mass integration of distinguishable technologies with varied purposes, into mobile devices.[5] This is also the case in other markets, including Android-based technologies and other smartphones across the market. The particular rise in popularity of Apple Computer Inc.'s iPhone is driven by the corporation's constant innovations and out-of-the-box utilities.[10] More recently, Apple Computer Inc.'s ambition to improve its presence led to the inclusion of their first-generation LiDAR (hereinafter referred to as iLiDAR) sensor incorporated into their new iPad Pro and iPhone 12 Pro models and successive devices beyond, up to the more recent iPhone 15 range.[10-12] The incorporation of the LiDAR sensor into the iPhone was intended to improve their measure application (app) capabilities by introducing depth sensing, portrait image capability, night mode performance and augmented and virtual video game functionality.[8] The iPhone 12 Pro and later models cost above (approximately) ZAR30 000, and now provide LiDAR capabilities to the public, marketed to provide real-time processing ability leading to comparable results, at no routine scanner maintenance fee. This package presents an incredible opportunity for a broader application base at the hands of

the market base. It is therefore scientifically interesting to benchmark these lower-cost laser scanners (e.g. iLiDAR) or sensors to engineering-grade technologies. Engineering-grade laser scanners are high-end specialised instruments known to obtain accurate spatial data effectively and accurately. However, their cost and application in less precise work is a barrier to adoption as they can range between ZAR600 000 and ZAR2 000 000, making them ideal investments for large projects and precise data programmes. Moreover, they require routine calibration and maintenance to ensure they satisfactorily deliver on accuracy over time.[6] Literature posits that there is insufficient documented knowledge on the iLiDAR capabilities, as there are very few studies that have captured the gains of this recent development from a scientific perspective.[13-15] The problem therefore lies in the limited amount of literature that benchmarks the emerging and low-cost LiDAR sensors to existing engineering-grade or commercial ones to inform multidisciplinary practice. Thus, this presents a knowledge gap for researchers to respond to, providing robust evidence that can assist users who may need to utilise mobile scan technology for diverse applications. This paper adopts a positivist paradigm and aims to test the accuracy of the iLiDAR sensor and its capabilities in generating a terrestrial laser scanner-derived point cloud by evaluating it against engineering-grade data collection sets. The research design is formulated to investigate the extent to which iLiDAR tools are available within the iPhone 12 Pro, compared to the engineering-grade laser scanner in terms of positioning, processing and visualisation. Monitoring how different technologies are being incorporated into mainstream technologies and how fast they grow in complexity will give an insight into the current accuracies and inform future data collection paradigms.

Associated theories and related literature

The measurement principles behind LiDAR scanning juxtapose the numerous benefits and applications of the technology, sharing a similar base concept of electronic distance measurements, which is now widely and at times passively used.[1] LiDAR at its core is a range detection method that uses a laser pulse to illuminate an object and measures the time taken for this pulse to return to the source, allowing the LiDAR scanner to accurately measure the distance between the sensor and the object.[3-5] From the interaction of the laser pulse with an environment, a three-dimensional (3D) impression of the real world is recreated with a collection of X, Y and Z coordinates for multiple locations. LiDAR sensing under low light conditions provides overall accuracies of 0.191 metres (m), 0.242 m and 0.345 m at 20 m, 40 m and 60 m altitudes.[6,7] Commercial terrestrial and mobile can achieve an accuracy of less than 20 millimetres (mm) for time-of-flight (TOF) scanners and less than 10 mm for phase difference scanners.[7] It should be noted that this is a very general estimate due to the wide range of LiDAR scanners available on the market, as many can provide between 3 mm and 6 mm accuracies. Mobile laser scanning takes the concept of terrestrial or ground-based laser scanning by adding real-time kinematic (RTK) GNSS and inertial measurement unit (IMU) systems for a moving platform.[2,7] This allows scanning to take place in rapid succession by driving a LiDAR scanner mounted to a vehicle/platform and producing a geo-referenced point cloud through registration.

Several models and frameworks have been postulated to explain user adoption of new technologies, with more than one theoretical approach required for a complete understanding of the broad trends we may see in technology uptake.[10-13] In this paper, technology adoption theories and models are not primarily significant given the research context. However, due to the practical nature of the findings in a rapidly advancing technological space, which may get dated quickly, there is merit in highlighting that technology adoption theories explain the changes and growth in the development of low-cost devices. In the case of the iLiDAR sensor, the literature articulates that third-party applications (apps) capitalised on the opportunity to use the LiDAR system in conjunction with the processing power of the iPhone's new bionic core, to provide 3D models just as the terrestrial LiDAR systems would.[14,15] The iLiDAR scanner is a combination of model sensors that provide users with the potential to approximate engineering-grade mapping capabilities, as it combines a refined GNSS receiver, enhanced gyroscope and accelerometer sensor as well as what is described as a high-precision camera sensor and LiDAR sensor.[16-19] The resulting 3D models then appealed to multiple disciplines who saw an opportunity to use the iLiDAR for reality capture and documentation. Thus, the additional ability to reconstruct 3D models using the iLiDAR is accredited not to Apple alone, but to several third-party sources who developed these apps to take advantage of the LiDAR system using the app developer's designed algorithm.[20] The sensor functions offer the same scientific concepts as their professional-grade counterparts albeit reduced to the most basic components. The apps can produce well-textured models using a simultaneous localisation and mapping (SLAM) algorithm, which in practice allows for the reconstruction of maps and models through the continuous updating of results using precise resection of the scanner's location and orientation.[21] SLAM is primarily used in mobile laser scanning, as the LiDAR scanner is mounted to some moving platforms or vehicles with the Global Navigation Satellite System (GNSS) and IMU systems constantly keeping track of the vehicle during acquisition. This would occur while the SLAM algorithm allows for registration and geo-referencing of all points in question.[21] Apart from its use in augmented reality and gaming, iLiDAR capabilities are promising for forensics, real estate, physics, archaeology and engineering documentation, as seen in a related study by Luetzenburg et al[14]. in which the accuracy of the iLiDAR systems was quoted to fluctuate between 3 cm and 6 cm.[6,14-17] This range demonstrates a high potential for the mapping of small-scale scenes where absolute positional accuracy may not be required, such as residential rooms and furniture in real estate.[14-16] However, despite these merits, the rapid growth to increase the reach of scan ability and the addition of specialised competencies to the selected mobile phones still demands immense focus to obtain more accurate measurements in real time, bearing in mind that engineering-grade scanners require recalibration to continuously provide quality data.[18,19,22]

Tavani et al.[15] describe the iPhone 12 Pro LiDAR system as a huge paradigm shift, improving the geospatial data acquisition process through acquiring 3D reconstruction models for fieldwork in real time. The study resonates with related works by other scholars and posits that such a capability in the hands of scientists, such as geologists, would improve research fieldwork opportunities by increasing access to low-cost LiDAR data and enhancing repeatability and transparency.[14-16] It is also noted as possible to add location detail or GNSS capability using the iPhone 12 Pro GNSS receiver; however, due to the low resolution available to these sensors, absolute accuracy would not approach engineering-grade standards. However, in a study conducted by Tamimi[17], it was concluded that the use of an external real-time kinematicGNSS receiver connected to the iPhone 12 via Bluetooth may provide much higher accuracy by using much more defined positional information. It is, therefore, one of the objectives of this research to evaluate whether the same is true without the external real-time kinematic receiver, particularly to evaluate if GNSS positions from the built-in receiver aid in any way to the final deliverable. In the same study by Tamimi[17], it was noted that the iLiDAR data accuracy did not increase too significantly between Generation 1 and Generation 2 LiDAR and camera systems. Its relative accuracy was exceptionally low when compared to that of the total station data or in comparison to what can be obtained with engineering-grade scanners, where most results attain less than 10 cm accuracy. According to the numerous studies conducted to compare two-point cloud data sets, including those of Tavani et al.[15] and Chauhan et al.[18], a trend for a defined accuracy assessment procedure that mirrors the one initiated by Parrish et al.[23] is evident. In the work conducted by Parrish et al.[23], the suitability of the iLiDAR sensor for forensic work was interrogated. The researchers used three techniques for LiDAR comparison, including a cloud-to-cloud (C2C) comparison, a rudimentary comparison of tape measurements, and a chalk outline clarity test. Similarly, Chauhan et al.[18] used point cloud comparison across two different registration algorithms. Individual point clouds were aligned together in Cloud Compare and made use of the C2C distance model computations. According to Chauhan et al.[18], 3D deviation analysis between point clouds is best performed using a C2C computation to provide for a comparison of the entire range of available points instead of a point-to-point or point-to-cloud method. In summary, the research cites that four distinct distance models can be used: the C2C comparison, C2C distance, multi-scale model-to-model (M2M) cloud comparison, and model-to-model cloud (M3C2) distance.

Conclusively and based on the literature[18,23,24] LiDAR scans conducted using two different scanning methods can be compared based on the overlap they provide, an estimation of the drift seen in the data and how the data sets manage elevation changes.[25,26] Such a comparison allows researchers to consider diverse data collection paradigms and evaluate how they conduct their work as seen in several studies.[27,28] The current study addresses its objectives by collecting data in the field and adopting a C2C and M3C2 approach in comparing the derived point clouds to take advantage of the statistical robustness of the comparison techniques which makes them robust and scientifically sound in tests.

Study area, materials and methods

The study site selected for this investigation was an old building area within Rondebosch, Cape Town, South Africa. Cape Town is on South Africa's southwestern coast close to the Cape of Good Hope and is the southernmost city on the African continent. The Snape and Menzies buildings[9], which date back to the early 1900s and are located in Rondebosch, Cape Town, were selected as study subjects as they presented regular but moderately complex facets with staircases and corridors that allow for robust testing for the case at hand. Another primary consideration towards the selection of the specific site and buildings was to opt for buildings that had been previously scanned with ease and therefore accommodate that learning to the iPhone 12 LiDAR scanner to mitigate any limitations it may face about the uncertainties in measuring highly complex structures as similarly proposed in related studies.[30] It is important to highlight for ethical considerations that the authors have obtained ethical clearance to conduct the work and have no conflict of interest or link whatsoever to Apple and its related stakeholders. The iPhone 12 Pro and its selection for the research was based on its availability to work, particularly as one of the few mainstream mobile technologies at present, which have introduced scanning technology. The researchers reflected on the overall research design to ensure that selected methods and data collection approaches would be optimal for the analysis. The initial fieldwork step was to set up ground control points (GCPs) of known geographic positions established using a highly accurate static GNSS approach as control data for the experiments. In establishing the control network, it was important to reflect on the intended provisional scan position to ensure that there would be enough control data in the system to cover the entire building with minimal error and an efficient geo-referencing of scans. Scanner setup positions were selected based on the amount of detail per area and the intent to introduce overlap between scans within proximity to the available ground control points. Two additional ground control points were placed outside the main test site and observed using virtual reference station real-time kinematic (VRS RTK) GNSS due to limited nearby control and a decline in satellite lock for GNSS reliability in densely built-up areas. As highlighted in the introduction, to evaluate the integrity of the iPhone 12 LiDAR scanner, it was to participate as would the engineering-grade scanners, and it was compared to both an X7 Trimble Laser scanner and a Z and F scanner. Using the iLiDAR scanner, the researchers were able to produce a workable deliverable by following a similar workflow to that of mobile laser scanning due to the similarities in the data acquisition process. To make room to register successive LiDAR point clouds, targets were placed on the walls wherever possible. These targets were black and white markers and were placed about 1.5 m above the ground. They were not placed throughout the site completely, because of weather conditions during fieldwork. In some places where markers could not be placed, noticeable features like the edges of clearly defined signs for the remaining segments of the building were identified and noted as useful. The iLiDAR, Z and F and the Trimble X7 (TLS) scanner were deployed, and point cloud data were collected. Because the iPhone LiDAR needed to be on the same coordinate system as the laser point clouds from engineering-grade scanners for comparison and to maintain good positional accuracy for the iPhone during data capture, open-source GNSS data were used. The iLiDAR data preparation included a registration step as the scanning process was done per wall to reduce strain on the iPhone processing unit. Each data bundle allowed for a C2C registration done in the field, where it was noted that the relationship between scans and ground control points was not as vital for registration and transformation, as in other scanning approaches. Thereafter, cleaning of the merged data bundles was done to remove all unnecessary features from the point cloud such that only the building features, paths and steps were accounted for. This was to ensure that the iLiDAR point cloud has similar details to the engineering grade data sets to enable a correct comparison between any two data sets. Once the data were ready, the phone scan data were compared to the engineering-grade point cloud to highlight similarities and key differences. An accuracy-based approach was adopted using algorithms within Cloud Compare using the C2C and M3C2 comparison.

Results and discussion

Using the field data for the iLiDAR scans, we followed a workflow like that of processing mobile laser scanning due to the similarities it holds with the iLiDAR data acquisition process. The cleaning of the collected data, particularly for iLiDaR, involved removing all unnecessary features from the point cloud such that only the building features, paths and steps were accounted for, as shown in Figure 1. This ensured that the precise point cloud had the same features as the engineering-grade LiDAR point cloud and would enable a correct comparison between these two data sets. During the importing process, the iLiDAR point clouds appeared to struggle to be interoperable with the software of choice, and when loaded, the point cloud would either be extremely small or disintegrate into an extensive line of points. A probable reason for this may be due to the processing done by the A14 64-bit ARMv8.5-A system on a bionic chip that may not be allowing the cloud to communicate to the computer correctly or may also deal with the way the iPhone 12 formed the geo-referenced file. This was as expected and seen with many Apple devices, where compatibility with other platforms may not always be smooth, as we see in related work by Allen et al.[27] and Liu et al.[30]

After cleaning, some further inspection of the data followed. Another challenge identified was that where the iLiDAR would only have minimal overlap areas at the edges for each wall, a least square adjustment matching solution would cause walls to be inverted in the opposite direction. This was attributed to the matching algorithm that views the maximum amount of area within the scan overlap region by tying in other similar features to one another. However, this caused the scans to be mirrored to one another so that the faces of the features overlapped

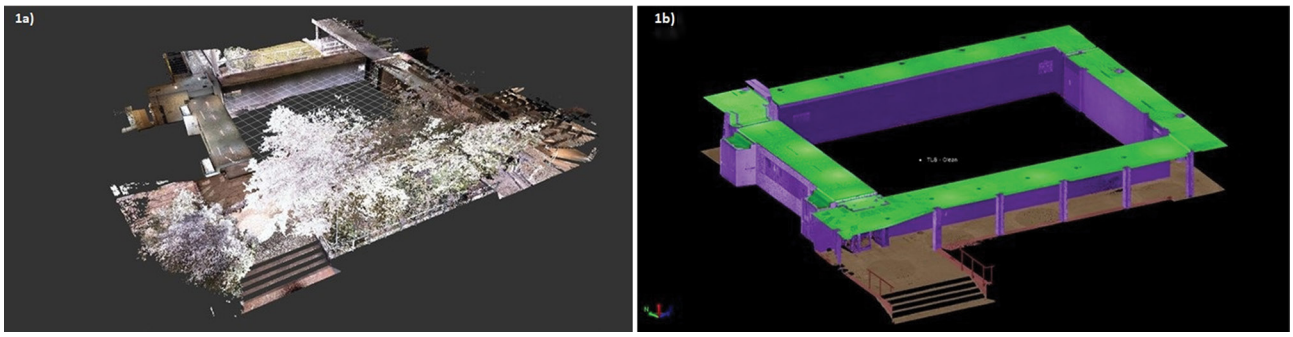


Figure 1: (a) Uncleaned Trimble laser scanner (TLS) point cloud; (b) cleaned and classified TLS point cloud.

on top of one another orientated in the same direction. This could be resolved with an additional point at the end of the wall to keep the orientation of each scan defined. Resolving this challenge meant that the iLiDAR scans now had to be aligned to the Trimble laser scanner and Z and F data set using an iterative closest point-based rotate or align function. It was finely aligned using the Finely Align function to correct for the remaining orientation discrepancies and overlay the two scans together. This effectively registered any two clouds together. In the case of TLS cloud, for example, this was achieved after 20 iterations and 5000 random point samplings (Figure 2).

The aligned scans merged into a one-point coherent cloud for both 2 m and 4 m scans. The 4 m scan contained the inside corridor. Because the inside corridor needed enough points to tie the scan into the remaining other scans, it was integrated into the 4 m scan. Although the scan could not have been done with a 4 m scan range, this choice is justifiable as scanning was done using the maximum possible distance away from the object. These merged scans were exported to Trimble Business Centre (TBC) to be classified and cleaned further. The accuracy of the iLiDAR point cloud could now be evaluated further by the researchers, using visual interpretation, C2C and M3C2 distance models as adopted from the literature. These compute metrics based on the distance between two respective points in a point cloud. The cloud distance methods, the detection and removal of outliers were facilitated using a Python programme. The results are summarised in Table 1 to highlight the change in centrality for each local model, providing insight into the skewness, concentration and distribution of the data. Visual analysis of these data was summarised in descriptive statistics tables which gave an overall mean and 95% confidence interval for each scan. The products of the M3C2 distance algorithms were also provided with their corresponding tabulations and heat maps concerning the C2C distance method. The M3C2 provided distance comparisons and summaries for its cloud distance computations and an analysis of the statistical models used for conveying the product's precision. Table 1 shows the descriptive statistics of each C2C local model distance in 2 m scans. The removal of outliers using percentiles caused a change in accuracy of 1.3 cm, indicating that these are groups of large values, up to 60 cm, in small proportions, implying that potential outliers were removed. Evaluating the data across all the local models revealed that the relative accuracy of the iLiDAR was between 6 cm and 8 cm, on average.

All models showed low variability with a standard deviation (SD) fluctuating between 5.5 and 7, and standard error measures close to zero, implying that these averages are good estimators of the true mean. In addition, the 95% confidence interval in each model allowed for a 3 mm window for its estimation of the true mean, implying strongly that these means closely approximate this value. The final outputs of each C2C distance local model used in the iLiDAR comparison for the 2 m scan were also prepared with its scalar field colour ramp, as illustrated in Figure 3, showing the colour corresponding to the distance calculated and a 10 m scale bar. The data set showed the visual distributions of the departures across the object surface, revealing the areas demonstrating the most and least variation from the TLS data set. Red regions on heat maps remained consistent between all local models; however, some models are more lenient with reporting the effect these areas have on the data. The red areas within the nearest neighbour (NN) and two-dimensional triangulation (2DT) models have more missing data in these regions, which indicates that these are the locations where most outliers were removed. The converse also remains true regarding blue areas, showing very stable results across all models having very dense point counts and showing extraordinarily slight variation across all models, such as the west wall. Areas of interest regarding larger error values include the north (front-facing wall) entrance and its adjacent wall segments. These areas of interest appear more speckled in the least squares plane (LSP) and quadratics height function (QHF) models with very random error responses ranging from exceptionally large, 22 cm, and exceedingly small, close to zero.

From the results of the above methodology, we aimed to evaluate whether the factors of our initial view on the comparison of accuracies of iLiDAR with commercial scanners would coincide with what we observed in the field. A C2C and M3C2 distance assessment provided an approximation of the iLiDAR accuracy using descriptive statistics providing different averages for the distance discrepancies, giving a general idea of what the system can give under a 95% confidence. In addition, the use of the root-mean-square (RMS) and Chi-squared results gives a final estimate of the average error observed and shows how well the data can be modelled. The results provided a comparison of distances across the 2 m and 4 m data sets to evaluate if the scanning method yields contrary results to the view or reasserts them. Upon visual inspection of the cloud data sets, all point clouds showed significant departures and large segments of discontinuity in the iLiDAR clouds. However, as a general summary, the iLiDAR seemed to approximate the TLS data set well, specifically within the 4 m scan of the data in the negative direction and represented points behind the wall which could not have been possible. This gave more credence that the lenient local models that produce more noise are utilising erroneous inclusions in their computations. For each local model, the data conveyed that the best estimate of the mean using that algorithm is only 1 mm different from the true mean. This did not imply that the population mean for the iLiDAR scan was 1 mm away from these estimates, but it reported on the confidence we have in the mean computed for that specific algorithm. As the 4 m scan outperformed the 2 m in this manner, it implied that the 4 m provides the more authentic estimation of the accuracy available to the iLiDAR. This was further seen by considering the descriptive statistics tables to their Weibull distribution, where beta values, *b*, decreased from about 0.8–0.6 to 0.33–0.26, which meant that there was more conformity in terms of a lack of reliability to reach the larger error values in the 4 m scan than in the 2 m. Within the 2 m scan, it was noted that due to this, the noisier local models had to come into further evaluation because it was known that it is wrong to assume lower reliability to reach larger values in these areas as we know that the iLiDAR skewed because of drift. However, in the 4 m scan, there was more agreement between the models, and although the reliability on achieving larger values decreased, there was more trust in this assessment and we observed fewer drift errors (see Figure 4).

In addition, because of the scale and shift parameters, it was seen that the reliability was more defined over its error classes, which implies that the Weibull distribution is reporting on the full scope of the errors possible and not understating the effect of the larger errors. It must also be noted that the minimum value for each local model is not exactly zero, meaning

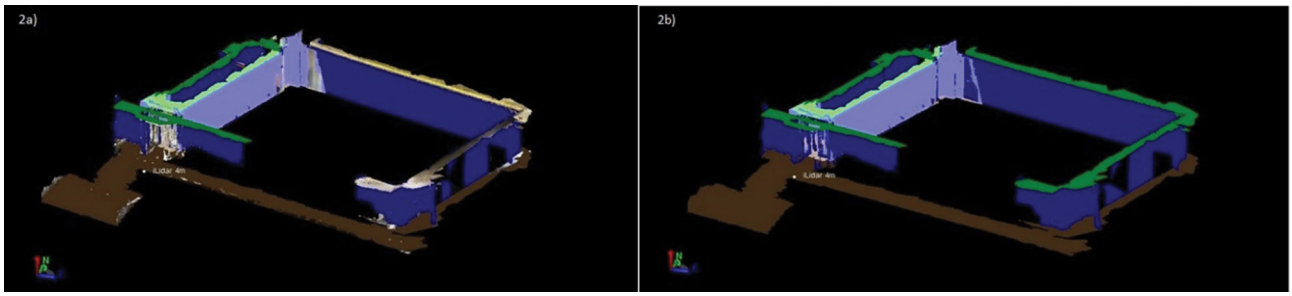


Figure 2: Images of (a) iLiDAR and (b) TLS point cloud alignment and cloud registration in Cloud Compare for 4 m scan.

Table 1: Summary statistics of C2C distance local model (iLiDAR 2 m)

Nearest neighbour		Least squares plan	E	2D1/2 Triangulation	N	Quadratic height function	
Mean (m)	0.080	Mean (m)	0.060	Mean (m)	0.080	Mean (m)	0.061
Standard error (m)	0.00004	Standard error (m)	0.00003	Standard error (m)	0.00004	Standard error (m)	0.00003
Mode (m)	0.003	Mode (m)	0.001	Mode (m)	0.003	Mode (m)	0.002
Median (m)	0.060	Median (m)	0.043	Median (m)	0.060	Median (m)	0.041
Standard deviation (m)	0.070	Standard deviation (m)	0.055	Standard deviation (m)	0.070	Standard deviation (m)	0.057
Sample variance (m)	0.005	Sample variance (m)	0.003	Sample variance (m)	0.005	Sample variance (m)	0.003
Range (m)	0.314	Range (m)	0.228	Range (m)	0.314	Range (m)	0.233
Maximum (m)	0.314	Maximum (m)	0.228	Maximum (m)	0.314	Maximum (m)	0.233
Minimum (m)	0.000	Minimum (m)	0.000	Minimum (m)	0.000	Minimum (m)	0.000
Points	3 131 245	Points	3 165 355	Points	3 129 841	Points	3 160 523
Sum (m)	250 782.6	Sum (m)	190 980.5	Sum (m)	250 440.5	Sum (m)	191 504.1
Classes	1770	Classes	1780	Classes	1770	Classes	1778
Confidence interval (CI) (95%)							
Lower CI (m)	0.079	Lower CI (m)	0.059	Lower CI (m)	0.078	Lower CI (m)	0.059
Upper CI (m)	0.082	Upper CI (m)	0.062	Upper CI (m)	0.082	Upper CI (m)	0.062

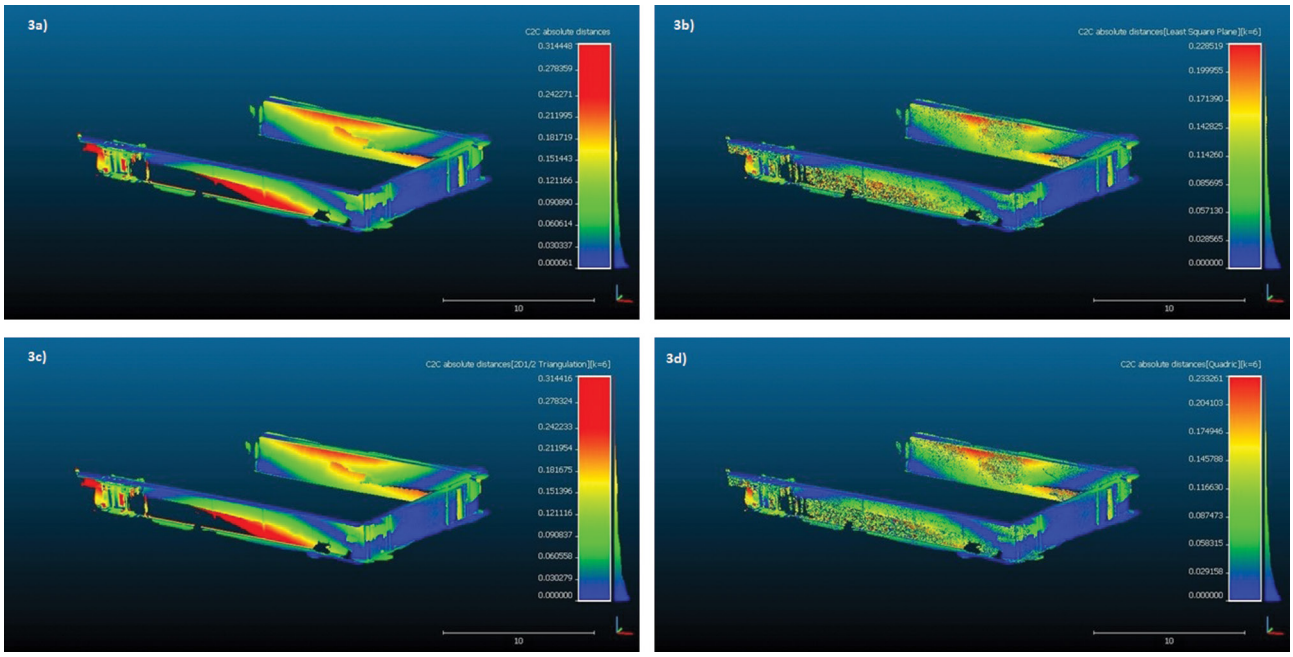


Figure 3: (a) C2C distance – nearest neighbour for 2 m scan; (b) C2C distance – least squares plane for 2 m scan; (c) C2C distance – two-dimensional (2D) half (½) triangulation for 2m scan; (d) C2C distance – quadratic height function for 2 m scan with its scalar field ramp, scale bar and orthogonal axes.

that the iLiDAR is not precisely synchronous with the TLS data set but that based on the local model used, it very closely approximates our TLS in these areas. This was seen as we expanded the values to further decimal places and saw the residual error in the iLiDAR measurements. However, these small errors were sub-millimetre and are not measurable to an exact value in practice. Based on data, long scan lines only increased the chance of misalignment due to drift. A possible reason for this relation between the scan length and misalignment was due to a decreased potential for overlap during the scanning process as longer scan lines make the iLiDAR scan more dependent on maintaining good IMU capability, that is, longer scan lines needed a very good fix on its orientation and position in space than shorter scans, in addition to less available area for overlap. It must be carefully noted that it is a lack of overlap in conjunction with the limited IMU ability of the iPhone that produces these errors. This is reciprocated in the areas within the 4 m scan which had shorter scan lines but much more stable deliverables. Because a 4 m scan will have a larger scope of the object being scanned, there will be a greater opportunity for overlap. This was therefore the primary reason that the 2 m scan failed to reach higher accuracies in comparison to the 4 m scan. However, the iLiDAR point accuracy remains satisfying at < 1 cm at points close to the start as advertised, as there still exists the same drift as the iPhone 12, which makes it incompatible and insufficient for mapping and even dangerous. To that point, however, we can supplement the iPhone data with correctly surveyed control points; we can reduce this error such that one could

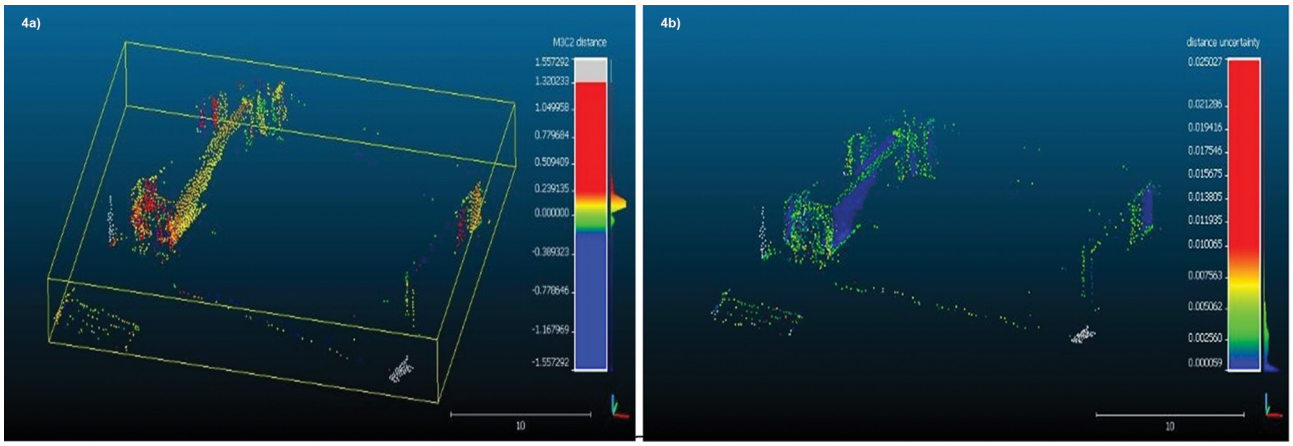


Figure 4: Illustration of where differences or errors lay in the 2 m scan labelled (a) and 4 m labelled (b) (blue region shows the M3C2 distance, low uncertainty ranging through green, yellow, and orange, while red regions are high uncertainty and high M3C2 distances).

say it is relatively accurate if we understand what the results should be. The iPhone 13, a later model than the iPhone 12, does hold its positional accuracy overall, and its colourised points line up neatly with the point cloud.

Conclusions and recommendations

This study aimed to test the accuracy of the iLiDAR sensor and its capabilities in generating a terrestrial laser scanner-derived point cloud by benchmarking it to a terrestrial scanner cloud. The above results indicate that the iLiDAR performed well in the context of a fit-for-purpose tool as seen in the 4 m scans. The same can be observed in the 2 m scans. However, it can be noted that for positional accuracy in cases where highly accurate positional detail is of importance, its inbuilt GNSS capability struggled to provide adequate absolute accuracy as anticipated. This did not aid much, if at all, in maintaining good iLiDAR measurement capability, much as has been seen in mobile laser scanning. To address this, mitigation methods of providing optimal results from the iLiDAR system strongly recommended that a proper stabiliser be used for the acquisition of the iLiDAR if greater accuracy is desired. This would allow the IMU capabilities of the iPhone to work optimally with the SLAM algorithm, and, in addition, the set-up may also benefit from connecting an external GNSS receiver as seen in Tamimi[17]. We also recommend further research towards more integrated approaches with structures from motion photogrammetry to deliver textured models to users and reinforce the limitations of the iLiDAR system which may provide the near 1 cm accuracy cited by Apple developers and Luetzenburg et al.[14].

Acknowledgements

We acknowledge Amahle Mkhize, Adnaan Parker and Suleiga Schroder for field data support.

Data availability

The data supporting the results of this study are available upon request to the corresponding author.

Declarations

We have no competing interests to declare. We declare that artificial intelligence (AI) was not utilised in the study or in the writing of the manuscript.

Authors' contributions

K.J.J.: Conceptualisation, data collection, data analysis, writing – the initial draft. M.S.: Project leadership, student supervision, writing – the initial draft, writing – revisions. Both authors read and approved the final manuscript.

References

1. Wang W, Zhao W, Huang L, Vimarlund V, Wang Z. Applications of terrestrial laser scanning for tunnels: A review. J Traffic Transp Eng (English Edition). 2014;1(5):325–337. https://doi.org/10.1016/S2095-7564(15)30279-8
2. Pajares G. Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs). Photogramm Eng Remote Sens. 2015;81(4):281–330. https://doi.org/10.14358/PERS.81.4.281
3. Shen N, Wang B, Ma H, Zhao X, Zhou Y, Zhang Z, et al. A review of terrestrial laser scanning (TLS)-based technologies for deformation monitoring in engineering. Measurement. 2023;223, Art. #113684. https://doi.org/10.1016/j.measurement.2023.113684
4. Li X, Liu C, Wang Z, Xie X, Li D, Xu L. Airborne LiDAR: State-of-the-art of system design, technology, and application. Meas Sci Technol. 2020;32(3), Art. #032002. https://doi.org/10.1088/1361-6501/abc867
5. Opitz RS, Ryzewski K, Cherry JF, Moloney B. Using airborne LiDAR survey to explore historic-era archaeological landscapes of Montserrat in the Eastern Caribbean. J Field Archaeol. 2015;40(5):523–541. https://doi.org/10.1179/2042458215Y.0000000016
6. Schulz T. Calibration of a terrestrial laser scanner for engineering geodesy [doctoral thesis]. Zürich: Institute of Geodesy and Photogrammetry; 2007. p.160. https://doi.org/10.3929/ethz-a-005368245
7. Ortenberg F. Hyperspectral sensor characteristics: Airborne, spaceborne, hand-held, and truck-mounted; integration of hyperspectral data with Lidar. In: Thenkabail PS, Lyon JG, Huete A, editors. Fundamentals, sensor systems, spectral libraries, and data mining for vegetation. Boca Raton, FL: CRC Press; 2018. https://doi.org/10.1201/9781315164151-2
8. Ventola CL. Mobile devices and apps for health care professionals' uses and benefits. Pharm Ther. 2014;39(5):356–364. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4029126/
9. Jaboyedoff M, Oppikofer T, Abellán A, Derron MH, Loye A, Metzger R, et al. Use of LIDAR in landslide investigations: A review. Nat Hazards. 2012;61(1):5–28. https://doi.org/10.1007/s11069-010-9634-2
10. Tatsumi S, Yamaguchi K, Furuya N. Forest scanner: A mobile application for measuring and mapping trees with LiDAR-equipped iPhone and iPad. Methods Ecol Evol. 2023;14(7):1603–1609. https://doi.org/10.1111/2041-210X.13900
11. Gera R, Chadha P, Ahuja V. Mobile app usage and adoption: A literature review. Int J Electron Bus. 2020;15(2):160–195. https://doi.org/10.1504/IJEB.2020.106546
12. Islam MS, Grönlund Å. Factors influencing the adoption of mobile phones among the farmers in Bangladesh: Theories and practices. Int J Adv ICT Emerg Reg. 2011;4(1):4–14. https://doi.org/10.4038/icter.v4i1.4670
13. Lai PC. The literature review of technology adoption models and theories for the novelty technology. J Inf Syst Technol Manag. 2017;14:21–38. https://doi.org/10.4301/S1807-17752017000100002

14. Luetzenburg G, Kroon A, Bjørk AA. Evaluation of the Apple iPhone 12 Pro LiDAR for an application in geosciences. Sci Rep. 2021;11(1), Art. #22221. https://doi.org/10.1038/s41598-021-01763-9

15. Tavani S, Billi A, Corradetti A, Mercuri M, Bosman A, Cuffaro M, et al. Smartphone assisted fieldwork: Towards the digital transition of geoscience fieldwork using LiDAR-equipped iPhones. Earth-Sci Rev. 2022;227, Art. #103969. https://doi.org/10.1016/j.earscirev.2022.103969

16. King F, Kelly R, Fletcher CG. Evaluation of lidar-derived snow depth estimates from the iPhone 12 pro. IEEE Geosci Remote Sens Lett. 2022;19:1–5. https://doi.org/10.1109/LGRS.2022.3166665

17. Tamimi R. Relative accuracy found within iPhone data collection. Int Arch Photogramm Remote Sens Spatial Inf Sci - ISPRS Arch. 2022;43(B2-2022):303–308. https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-303-2022

18. Chauhan I, Rawat A, Chauhan MPS, Garg RD. Fusion of low-cost UAV point cloud with TLS point cloud for complete 3D visualisation of a building. In: Proceedings of the 2021 IEEE International India Geoscience and Remote Sensing Symposium (InGARSS). Ahmedabad: IEEE; 2021. p. 234–237. https://doi.org/10.1109/InGARSS51564.2021.9792104

19. Monsalve A, Yager EM, Tonina D. Evaluating Apple iPhone LiDAR measurements of topography and roughness elements in coarse bedded streams. J Ecohydraulics. 2023;1:1–11. https://doi.org/10.1080/24705357.2023.2204087

20. Apple Newsroom. Apple unveils new iPad Pro with breakthrough LiDAR Scanner and brings trackpad support to iPadOS [webpage on the Internet]. c2020 [cited 2024 Mar 01]. Available from: https://www.apple.com/newsroom/2020/03/apple-unveils-new-ipad-pro-with-lidar-scanner-and-trackpad-support-in-ipados/

21. Huang L. Review on LiDAR-based SLAM techniques. 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML); 2021 November 14; Stanford, CA, USA. IEEE; 2021. p. 163–168. https://doi.org/10.1109/CONF-SPML54095.2021.00040

22. Chase P, Clarke K, Hawkes A, Jabari S, Jakus J. Transforming construction with reality capture technologies. Fredericton: University of New Brunswick; 2022. https://doi.org/10.57922/tcrc.645

23. Parrish CE, Jeong I, Nowak RD, Smith RB. An empirical comparison of full waveformLidar algorithms. Photogramm Eng Remote Sens. 2011;77(8):825–838. https://doi.org/10.14358/PERS.77.8.825

24. Sepasgozar SM, Wang C, Shirowzhan S. Challenges and opportunities for implementation of laser scanners in building construction. 2016 Proceedings of the 33rd International Symposium on Automation and Robotics in Construction; Auburn, USA. The International Association for Automation and Robotics in Construction; 2016. p. 742–751. https://doi.org/10.22260/ISARC2016/0090

25. Ahmed Fuad N, Ismail Z, Majid Z, Darwin N, Ariff MFM, Idris KM, et al. Accuracy evaluation of digital terrain model based on different flying altitudes and conditional of terrain using UAV LiDAR technology. IOP Conf Ser: Earth Environ Sci. 2018;169(1), Art. #012100. https://doi.org/10.1088/1755-1315/169/1/012100

26. Ahmed Fuad N, Yusoff AR, Ismail Z, Majid Z. International archives of the photogrammetry, remote sensing and spatial information sciences - ISPRS archives. 2018;42:11–21. https://doi.org/10.5194/isprs-archives-XLII-4-W9-11-2018

27. Allen S, Graupera V, Lundrigan L. Pro smartphone cross-platform development: iPhone, Blackberry, Windows mobile and Android development and distribution. New York: Apress; 2010. https://doi.org/10.1007/978-1-4302-2869-1

28. Mallet C, Bretar F. Full-waveform topographic Lidar: State-of-the-art. ISPRS J Photogramm Remote Sens. 2009;64(1):1–16. https://doi.org/10.1016/j.isprsjprs.2008.09.007

29. University of Cape Town. Buildings, departments and offices [webpage on the Internet]. c2024 [cited 2024 Mar 01]. Available from: https://uct.ac.za/contacts-maps/buildings-departments-and-offices

30. Liu C, Zhu Q, Holroyd KA, Seng EK. Status and trends of mobile-health applications for iOS devices: A developer's perspective. J Syst Softw. 2011;84(11):2022–2033. https://doi.org/10.1016/j.jss.2011.06.049

AUTHORS:
Tricia Harraway[1]
James Bekker[1]

AFFILIATION:
[1]Department of Industrial Engineering, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
James Bekker

EMAIL:
jb2@sun.ac.za

DATES:
Received: 06 Mar. 2024
Revised: 27 June 2024
Accepted: 12 Aug. 2024
Published: 04 Dec. 2024

HOW TO CITE:
Harraway T, Bekker J. Voucher tokenisation using blockchain and smart contracts to support people in need. S Afr J Sci. 2024;120(11/12), Art. #18150. https://doi.org/10.17159/sajs.2024/18150

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITORS:
Michael Inggs
Thywill Dzogbewu

KEYWORDS:
blockchain, digital vouchers, disintermediation, smart contracts

FUNDING:
None

Voucher tokenisation using blockchain and smart contracts to support people in need

In South Africa, many people are homeless or doing informal work for which they receive small amounts of cash from caring individuals. The world is moving towards cashless transactions, but devices are needed to support that move. Many people in need cannot afford such devices and usually receive cash, but fewer donors carry cash. Consequently, people in need receive less informal financial support. We propose a system that allows donors to give digital vouchers that can be redeemed at participating stores and institutions of care. This study aimed to investigate the use of blockchain technology in digital voucher management and to demonstrate the application of smart contracts to disintermediate the value transfer process specific to the donation process. A demonstrator was built to include a front end for the user to interact with and a back end containing the application logic, which was built on the Polygon blockchain, a second-layer solution for the Ethereum blockchain. The model included tokenising vouchers as non-fungible tokens, and the smart contracts governed their logic and the conditions to be met. The demonstrator was validated using smart contract and unit tests to evaluate the security and functionality. While the model was not implemented in reality, a fully functioning demonstrator was developed. The platform achieved the aim of disintermediating the voucher management process. A real-world implementation could help many in need to receive tokens for food, shelter and clothing from direct, individual donors.

Significance:

We developed an architecture for a digital voucher donation management system that enables individual donors to transfer cashless vouchers to recipients in need. The recipients can then exchange the vouchers for necessities such as shelter, food and clothing. Such a system will become essential in the near future as we move towards a cashless society. The proposed system is safe because it is based on blockchain technology.

Introduction

South Africa has a complicated history of racial inequality and is the world's most economically unequal country, according to the Gini coefficient. This coefficient is an index of economic inequality and ranges from 0 (perfect economic equality) to 1 (complete economic inequality). The value of the Gini coefficient for South Africa was 0.63 in 2023, which is the highest value in the world.[1] Furthermore, the top 10% of the South African population earn more than 65% of the national income. Various factors contribute to these figures, and financial exclusion is considered a key factor, according to the World Bank.[2] Financial inclusion (FI) is the state in which "all persons have timely and fair access to appropriate, fair, and affordable financial products and services", according to the South African Financial Sector Regulation Act.[3] The World Bank[4] asserts that this cannot be guaranteed in practice, particularly for lower- to middle-income countries. The significance stems from the fact that, as of 2018, research[4] shows that more than one billion people around the world lack a formal means of proving their identity, of which about 81% reside in South Asia and Sub-Saharan Africa. In the poorest 20% of countries, 45% of people lack a national identity document. Formal or legal identity documentation in South Africa is either in the form of a green bar-coded identity book, identity card or valid South African passport issued by the Department of Home Affairs.[5] In South Africa, individuals without formal identity documentation are severely disadvantaged and inhibited from participating economically and socially. They are likely to be excluded from access to basic services or opportunities such as opening a bank account, applying for a mobile service provider, using a money market, applying for formal employment or government housing and receiving various forms of social grants.[6] Economic inequality and education levels are usually linked, and many people are denied an education due to poverty.[7,8]

The South African Government has the developmental objective of closing the wealth and financial inclusion gap[9], and the National Development Plan for 2030 presents a roadmap describing the process to relieve the poverty and inequality[10] in the country. Citizens should, however, not leave it to their governments alone to improve people's lives. They can also help others as individuals. Claude Ake, a renowned Nigerian political philosopher, said, "It is the ordinary people who alone can make development sustainable, and development has not really occurred until it is sustainable. The people make development sustainable only insofar as its content becomes an integral part of their lives"[11]. Community philanthropy by civil participation is one way for individuals to uplift their peers.

In South Africa, philanthropy is multifaceted: it ranges from structured, formal giving mechanisms, such as corporations, organisations or high-net-worth individuals to informal peer-to-peer forms of supporting the underserved.[12] Informal philanthropy is usually done by donating money or time directly to the underserved, and not via an organisation.[13] In lieu of cash donations, the Western Cape government (2022) promotes the donation of food, clothing and household goods or vouchers. Informal donations are typically not premeditated acts of kindness. It is uncertain when or where a donor would get in contact with someone in need, and it is unlikely that a donor travels with donatable items at all times. As a result, the standard method of helping someone in need is to give money. This is only possible if the donor has cash on them. This is unlikely to remain the case in the future because South Africa's payment environment is gradually moving away from cash and towards digital methods.

According to the FinMark study[14], 77% of people in South Africa used digital services for payments, with 71% of those payments going towards receiving income. In this study, we examined the country's digital payment landscape as well as the use of and access to financial services.

A cashless society using e-wallets, credit cards, mobile money and cardless application (app) payment systems is predicted by the growth trajectory in digital payment usage.[15] According to the Better Than Cash Alliance[16], these payment methods are becoming more widely accepted in communities and can reduce the disparity in financial exclusion. This hypothesis is supported by BankservAfrica[17] which claims that innovative digital solutions for financial services targeted at the less fortunate can benefit societies. This also applies to donation platforms and charities that may use digital means to disperse their aid.

We propose a digital voucher management system to facilitate voucher donations to ensure that the recipient uses it for responsible purposes[18] to empower the less fortunate and make 'responsible giving' more convenient. The conservative assumption is that the beneficiary does not own a mobile device and will be issued with a simple, cheap device such as a quick response (QR) code. For this reason, digital vouchers and card-based systems could be considered as a digital means for distributing donations, and to enable this, we discussed the potential of blockchain technology (BCT) as a mechanism for informal peer-to-peer charity donations in South Africa. We demonstrated the concept with a digital voucher donation management system (DVDMS) developed using engineering methods and available software products. The demonstrator showcased the practical implementation of BCT and its ability to enable digital donations through voucher tokens while ensuring security and accountability for all parties involved in a donation.

The concept of the DVDMS is presented now, while the design will be presented later. Caring institutions such as night shelters, retailers and individuals create digital vouchers for products on the DVDMS. These products could be a loaf of bread, a can of milk, a pack of nappies or a night at a shelter. Donors consider the digital vouchers on offer and buy one or more. When the donor wants to give a digital voucher to a beneficiary, the donor uses a mobile device to scan the QR code of the beneficiary, thus reserving the specific product for them. The beneficiary presents the QR code to the vendor in exchange for the product. No cash is involved, and the beneficiary does not need any expensive devices. The beneficiary, when registering on the DVDMS, receives a near-field communication (NFC) card and a QR code for identification. All transactions are secured and visible on a blockchain platform.

Voucher systems

Vouchers and voucher systems have a long history, having been used as money or as a means of redeeming goods and services in ancient civilisations.[19] Vouchers have been used for numerous purposes in the modern era, such as exchanging goods between businesses, marketing for drawing in clients and distributing and redeeming government benefits. Vouchers may be used as gift certificates or as part of a loyalty programme, as a way to redeem a stay at a hotel or a meal at a restaurant or to redeem a ticket or pass for travel.[20] Vouchers serve to transfer value without giving up the tangible item up front, which makes them useful for charitable donations and situations where the donor might not have cash on hand and/or the beneficiary does not have a bank account. Additionally, vouchers can stimulate the local economy by supporting small businesses and promoting consumer spending.[21]

The introduction of digital vouchers is one of the many consequences of the increased use of digital technologies where the need for digital solutions to traditional methods arose. Digital vouchers provide a convenient and secure means of exchanging value and goods. In recent years, consumers have been adopting digital vouchers more and more as a method of payment instead of traditional cash or credit card options.[22]

Decentralised donation voucher systems involve the use of digital vouchers that are distributed among a community of donors. These vouchers can then be redeemed for goods or services by recipients such as vulnerable individuals or organisations. By using a decentralised platform, these donation systems can promote horizontal community philanthropy, which is based on the idea of collective action and shared responsibility for the well-being of the community. Next, we present a secure digital platform allowing the exchange of digital vouchers for individual charity.

Blockchain and vouchers

The blockchain concept was first introduced in a 2008 white paper by the pseudonymous individual or group known as 'Satoshi Nakamoto'.[23] BCT has recently gained attention as a solution for various industries. It is a decentralised and distributed digital ledger used to record transactions across a network of computers. It started as a secured transactional concept and is a distributed ledger technology (DLT) that allows multiple parties to access and update a shared database without needing a central authority.[24] Transactions recorded on a blockchain are grouped into blocks and linked in a chronological chain, creating an immutable record of all transactions on the network. Transactions are validated by consensus among the nodes.[25] Decentralisation is a type of network in which nodes are independent of a central point or master node, and rather share a distributed control among the other nodes on the network.[25] It makes blockchains more resilient to attacks and failures because there is no single point of failure in the network. BCT presents promising tools for various applications, including secure and transparent record-keeping, identity verification and decentralised financial systems. One well-known application of BCT is cryptocurrency, particularly Bitcoin, which utilises the technology as a decentralised record-keeping database for monetary transactions.[23]

BCT offers smart contract functionality.[26] Smart contracts are digital contracts that hold their logic in code, which is automatically executed by a computer program, whereas a traditional contract is a legal agreement that is manually executed. Smart contracts can be used to create and manage vouchers, ensuring that the terms of the voucher are met before it can be redeemed. For example, a smart contract could be set up to allow a voucher to be redeemed only after a certain date, or only if certain conditions are met. Smart contracts can also be used to track the usage of vouchers, providing transparency and traceability for both the issuer and the holder. The smart contract can also include a unique identifier that verifies the authenticity of the voucher.[27-29]

Blockchain contracts enforce their terms without reliance on a third party and create tamper-proof records due to the cryptographic and immutable nature of BCT.[30] Along with these programmable contracts, BCT facilitates the creation of non-fungible tokens (NFTs) that resemble unique, indivisible and in-exchangeable (for other tokens) digital assets.[31] On blockchain markets, smart contracts are often used to facilitate the buying and selling of NFTs. The funds are held by the smart contract, which functions as a digital escrow, until the buyer receives the NFT. This contributes to ensuring the security of the transaction and compliance with the terms of sale. NFTs and digital vouchers are similar in that they are both digital assets that can be bought, sold and traded. However, NFTs are unique digital assets that are verified on a blockchain, while digital vouchers are typically digital versions of physical vouchers that can be redeemed for goods or services. By using BCT, vouchers in the form of NFTs provide a new level of security and authenticity for both the issuer and the holder.

One of the biggest challenges with vouchers is ensuring their validity and preventing fraud.[31] BCT can help solve this problem by providing a secure and transparent way to create and track vouchers. Blockchain-generated digital tokens can be redeemed like coupons or gift cards. The tokens can be fungible or non-fungible, and a variety of token formats can be created, including the non-fungible ERC-721 tokens on the Ethereum blockchain. These tokens can be programmed to have specific attributes, for example, expiration dates, and can be tracked on the blockchain to verify their validity, just like a traditional voucher.

BCT can also be used for vouchers by creating decentralised applications (dApps).[25] dApps are digital applications that run on a blockchain network, allowing for a decentralised and trustless platform. dApps can be used to create and manage vouchers, enabling users to redeem vouchers directly on the blockchain without needing a centralised

intermediary. It can facilitate the management of charitable donations (in this case, a voucher token) by creating a secure and transparent record of an asset's ownership and location, and the conditions of the individual voucher donations made in a system.

This research aims to explore three use cases in the form of (1) vouchers as tokens, governed by the (2) logic of smart contracts, all on (3) a decentralised platform such as a blockchain. By building the DVDMS on a blockchain, it effectively disintermediates the distribution of voucher donations, resulting in a more direct form of transferring vouchers from those who have the capacity to give, and those who are in need. The smart contracts act as self-executing programs that execute the creation, purchase transfer (from vendor to donor), donation transfer (donor to beneficiary) and redemption transfer (donor to vendor), based on preset conditions.

System functional requirements

We present the functional requirements we anticipated a DVDMS should have. Before we do that, we explain the workings of the system in simple terms. Three role players enable the system operations. These are the beneficiary (person receiving a donation), donor (an individual wishing to donate to a beneficiary via a token) and vendor (usually a business entity that provides items that are associated with tokens). The vendor digitalises their products by creating voucher tokens that represent the value of their sales items. The donor buys a token from the vendor via the DVDMS platform; for example, a box of rusks. The donor then hands the token to the beneficiary, who will trade the token at the vendor for the box of rusks. Our vision is that the tokens will represent at least food items, clothing and shelter. The functional requirements for such a concept are shown in Table 1.

These requirements were used to develop a demonstrator, which will be discussed later.

Assumptions made

Assumptions were made prior to the design of the system to determine boundaries. These were:

1. The vendor is the end user and initiates the flow of information. We assume they have access to the Internet and computers.
2. The beneficiary is conservatively defined as someone who is not likely to have a cellphone or formal documentation, but may have access to a device via others, as well as having the capacity to carry a QR code or NFC chip device with them.
3. The custodianship and digitalisation process (i.e. assigning serial numbers representative of the token ID to the physical asset) of the vendor's products is not included in the scope.
4. Vetting the vendors is important, and it will still be necessary to implement operational protocols if the system is deployed. However, the system will provide the platform for the managing and transferring of the voucher token, not the onboarding of the vendors.
5. The beneficiary onboarding is not included in the system scope; however, it will be necessary to provide an in-person registration process for the beneficiary to receive their QR code and NFC card.
6. The onboarding payment gateway between fiat currency and cryptocurrency is not included.

System specification

We describe the DVDMS specification in words, as follows:

1. The DVDMS is a digital voucher donation management process built as a web-based platform. It enables the purchase and donation of voucher tokens, exchangeable for physical assets such as food, clothing and shelter.
2. Three different groups of stakeholders will be interacting on the DVDMS platform, namely the vendors, beneficiaries and donors.
3. Vendors will use a 'token template' to create vouchers for their products. The DVDMS will mint the vouchers as tokens, allowing the traceability of their movements on the blockchain.
4. The DVDMS is interoperable and has a self-sovereign identity service – this will facilitate the secure registration process of each of the stakeholders in the system and assign role-based access rights.

Table 1: Functional requirements (FRs) of the digital voucher donation management system (DVDMS)

FR ID	Functional requirement	Description
FR1	Connect digital wallet	System must enable the user to connect their wallet to the DVDMS via prompts or notification
FR2	Registration	System must facilitate the vendor's registration process
FR3	Log-in	System must automatically recognise the address of the user once their wallet is connected
FR4	Access control for vendor registration	System must grant registered vendors access rights to create (mint and list) a voucher token
FR5	Voucher platform catalogue	System must display the available vouchers for sale to the public in the form of a voucher catalogue
FR6	Manage voucher controls	System must allow the registered vendor to populate the voucher template and delete a voucher token
FR7	Manage voucher sale	System must list the vouchers for sale once the voucher has been minted as a token
FR8	Mint a voucher as a token	System must mint the vouchers as tokens once the vendor has populated the token template with the required attributes
FR9	Purchase, transfer and store tokens	System must create a marketplace that allows the donors to purchase and transfer the tokens from the vendor
FR10	Provide transparency and traceability of the tokens	System must provide full transparency of the movement of the tokens as they change ownership
FR11	Scan QR code wallet addresses	System must have a built-in QR code scanner that reads the address belonging to the code
FR12	Validate voucher	System must be able to validate the voucher by querying the blockchain and determining its status
FR13	Redeem voucher monetary value	System must allow the vendor to redeem the monetary value of the funds once the voucher has been successfully redeemed
FR14	Escrow function to hold funds	System must include an escrow service that automatically executes the release of the funds once the vendor proves the voucher has been redeemed

5. The DVDMS integrates the use of a digital wallet called MetaMask that will allow the donors to purchase the vouchers with their digital currency.
6. Voucher donations and transactions will be transparent and visible to all users via a dashboard.
7. The DVDMS allows the donor to donate directly to the beneficiary without any intermediary involvement.
8. Emphasis is placed on the ability of beneficiaries to participate (i.e. redeem vouchers and sign transactions) without the need for expensive technology such as a mobile phone.

The stakeholders

The three stakeholders are briefly discussed. Their role and technology requirements are emphasised.

1. Donor: The donor will exchange fiat currency (in this case, the South African rand) for tokens representing digital vouchers and will directly donate to the beneficiary of their choosing. The donor must have a mobile device capable of taking photos and scanning a QR code. In addition, the donor needs to have a digital wallet installed; for example, MetaMask.
2. Vendor: The vendor is a participant registered on the system by a system administrator and is offering something for sale on the platform. The vendors create the digital vouchers based on their products or services, and they define the conditions and intrinsic value of the digital voucher as well as the process of redeeming the value of the digital voucher. The vendor is the owner of the physical asset associated with the voucher donation token and plays an essential role as the redeemed value of the token needs to be guaranteed. The vendor needs a mobile device which is also capable of taking photos and scanning a QR code and which has a digital wallet installed, for example, MetaMask. A QR code scanner can be used in place of a mobile device; however, this requires additional integration procedures. An NFC reader is required to allow registered beneficiaries to sign their transactions by tapping their cards.
3. Beneficiary: The beneficiary is registered on the system, and they redeem the voucher for either goods or a service. Their interaction with the DVDMS is via a QR code assigned to them at registration and representing their digital wallet address. This allows the beneficiary to view their wallet contents as well as redeem the intrinsic value of the voucher at a participating vendor. The beneficiary needs only a QR code and an NFC card. The QR code represents their public key on the blockchain, similar to the use of a 'username' for an account, while the NFC card holds the encrypted private key, representing the secret password.

The information flow design is presented next, based on the above system specification and stakeholder description.

Information flow design

The registration of a beneficiary is shown in Figure 1. It is assumed that the beneficiary does not have a mobile device, but the option is still provided for. The beneficiary will register at a public place such as a clinic, participating retailer or other caring institution such as a night shelter. The NFC card and QR codes are inexpensive and safe to carry and use.

An activity diagram was constructed to explain the information flow and activities of the DVDMS and is shown in Figure 2. As mentioned before, there are three stakeholders in the DVDMS: the vendor, donor and beneficiary. The vendor agrees to participate in the system and provides products such as food items or shelter. The donor chooses an item from a selected vendor to donate and purchases a digital voucher for that item. The donor hands the beneficiary a QR code for the digital voucher. The beneficiary can then claim the item from the vendor, and the digital voucher expires. All transactions are securely handled by the blockchain platform, and the expiry of the voucher token is based on the conditions set in the smart contracts.

The life cycle and state transition of a digital voucher or token as used in the DVDMS are shown in Figure 3. The digital voucher is created by the vendor and the donor purchases it. An escrow function keeps the money paid safe while the digital voucher is transferred to the beneficiary. Once the digital voucher is claimed, it ceases to exist.

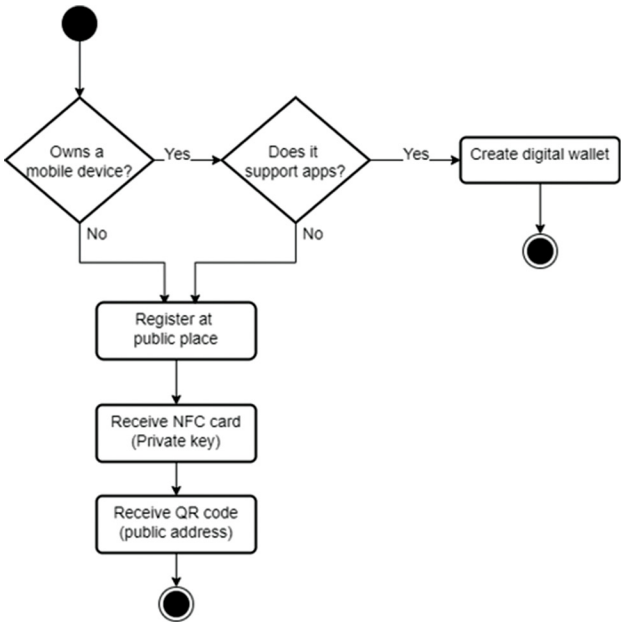


Figure 1: Beneficiary registration process.

A prototype to demonstrate the concept

A prototype was developed to conceptually demonstrate the use case of BCT as the foundation for a digital voucher management system. The blockchain-based DVDMS is abstracted as a front-end and back-end web application that facilitates the creation, purchase and redemption of digital voucher tokens for physical assets. The prototype was built using the following available technologies and components; the numbers in the list below correspond to the numbered labels in Figure 4:

1. The web application, which is accessible via any web browser. This is the front-end user interface of the DVDMS.
2. The donor or vendor, who would need to have a digital wallet to hold their cryptocurrency to integrate into the DVDMS. The donor needs to ensure that they have cryptocurrency in their digital wallet, while the vendor needs to ensure that they can receive cryptocurrency.
3. Authentication – To use the DVDMS, the user will need to authenticate transactions as well as user access rights using their digital signature from their cryptocurrency wallet. This is a cryptographic signature that authenticates the user as a donor, rather than a vendor, allowing them to purchase a digital voucher token, but not create one on the platform. The user will use their private keys that are stored in their digital wallet to prove to the platform that they control the respective wallet that is trying to 'connect' to the platform.
4. The MetaMask wallet, which is a cryptocurrency wallet, is the link between the user and the blockchain. This digital wallet allows a user to connect to the platform and interact with it by holding cryptocurrency to transact.
5. A gateway is required between the DVDMS and the InterPlanetary File System (IPFS), to make the IPFS service HTTP accessible from a non-decentralised Internet browser. This allows the DVDMS to upload and store files without needing a decentralised browser. The Infura gateway is used to support write access to IPFS.
6. Infura provides the tools and infrastructure that allow developers to take their ready blockchain application from testing to (scaled) deployment. Included in the development suite is access to IPFS and Polygon networks.
7. IPFS is founded on the concepts of peer-to-peer networking and content addressing and is a modular suite of protocols for

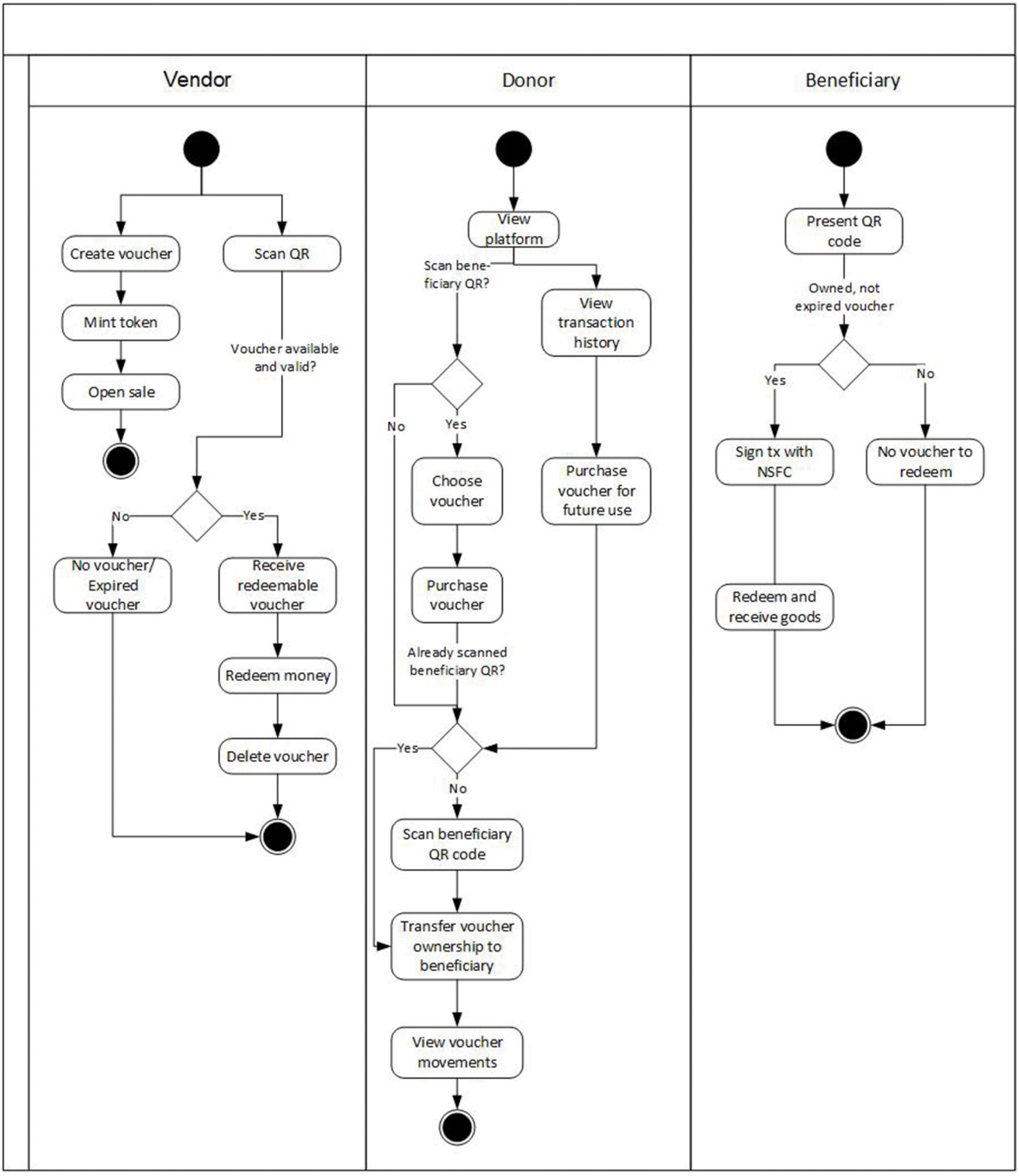


Figure 2: The digital voucher donation management system (DVDMS) activity diagram.

data organisation and transfer. This is used in a similar manner to a database; however, it is a decentralised storage service that stores data across multiple nodes on the network. The information associated with the digital voucher tokens is stored using this service, which also results in an entirely decentralised application.

8. The DVDMS host platform was built on the Polygon blockchain as it is a more affordable alternative to Ethereum. Polygon also allows for scaling of the network, and the resources for developers make it a more attractive blockchain to use for prototyping. Additional libraries, frameworks and tools are used to extend the capabilities of the Polygon blockchain. Hardhat is a development environment that allows for the debugging, testing and deployment of smart contracts on a test blockchain. Ganache enables local testing of smart contracts. OpenZeppelin is an open-source library for smart contracts, allowing developers to build on top of a secure foundation of code based on community-vetted security standards.

9. PolygonScan provides a blockchain explorer, search, API and analytics platform for Polygon. This was used to track any updates, changes or transactions specifically associated with the DVDMS platform. It was used to demonstrate the transparency and traceability of transactions on the blockchain, smart contract interactions and any digital voucher token movements on the Polygon network. It was also used to verify transactions and monitor the flow of assets within the application securely.

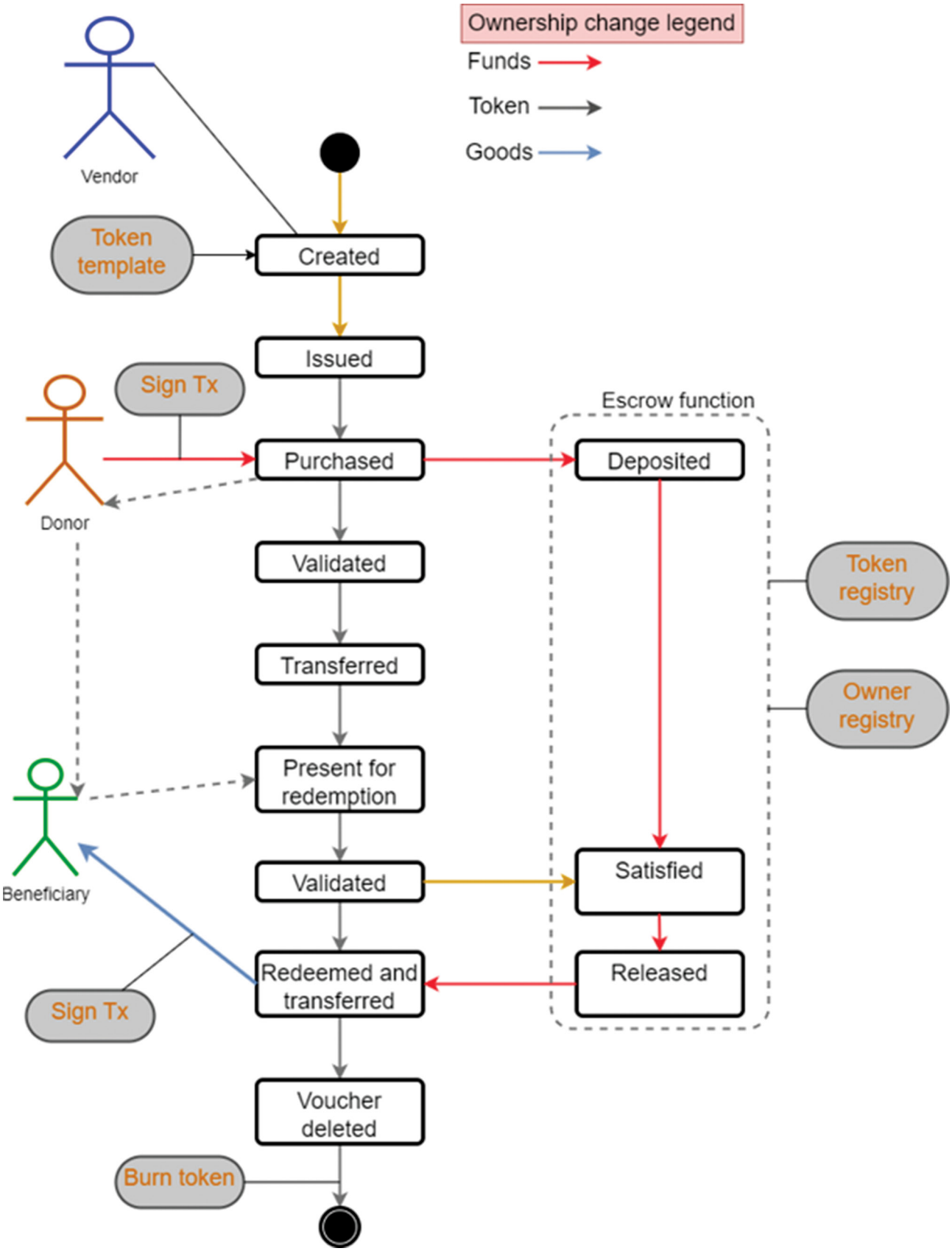


Figure 3: The life cycle and state transition of the voucher.

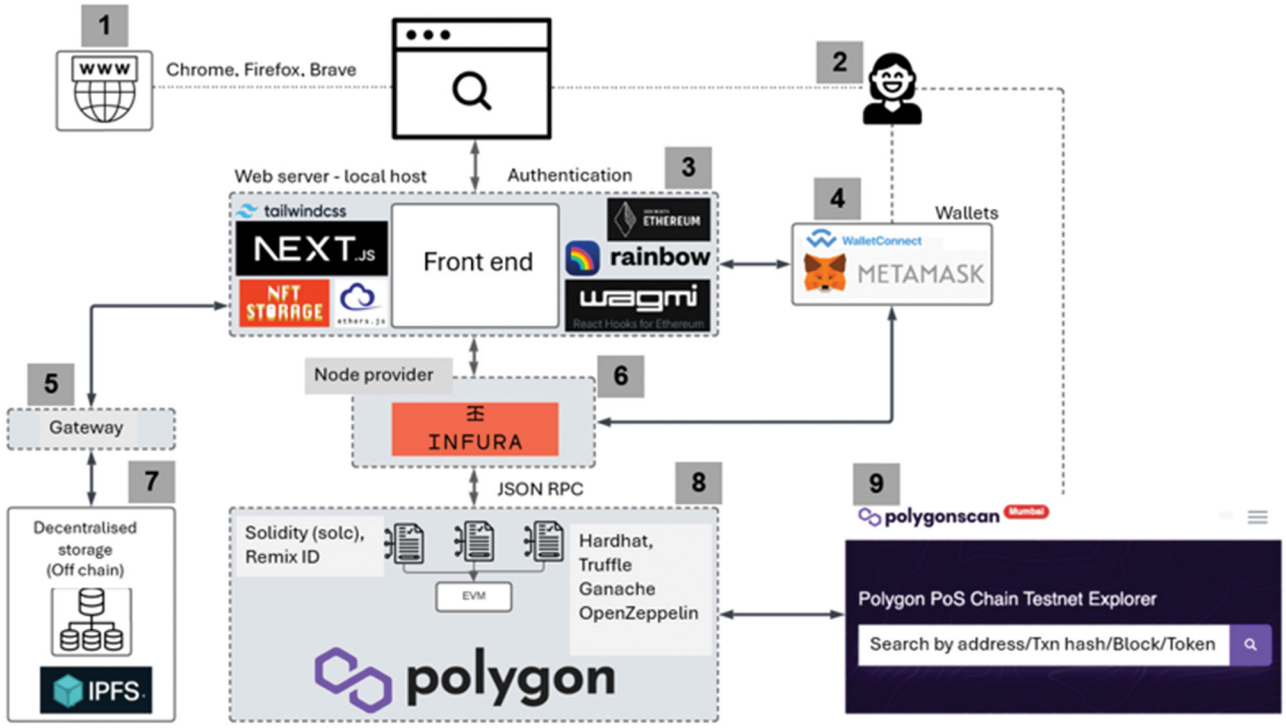


Figure 4: Digital voucher donation management system (DVDMS) software technology architecture.

Conclusion

We have proposed a system that will enable donors to give tokens to recipients with no cash involved. The donor, therefore, does not need to carry cash to support an underprivileged recipient, and the latter can safely redeem tokens from participating vendors. BCT has the potential to revolutionise the way digital vouchers are created and managed and was the cornerstone technology used to enable the proposed DVDMS. It enables the creation and tracking of digital vouchers on a safe platform that guarantees the vouchers' usefulness and validity. Furthermore, redeeming digital vouchers is automated using decentralised applications and smart contracts, improving efficiency and cost-effectiveness. As BCT develops, more companies can use it for financial instruments such as digital vouchers. Using BCT in a DVDMS has some limitations. The decentralised nature of BCT may pose regulatory challenges because it is unclear how existing regulations and laws apply to this new technology. Despite these limitations, the potential advantages of using BCT in a DVDMS make it an exciting area of research and development. A peer-to-peer DVDMS enabled by Blockchain could be a revolutionary way to manage charitable donations by providing transparency, security and efficiency to the donation process.

Data availability

There are no data pertaining to this article.

Declarations

This paper is based on the MEng thesis of T.H.[32] We used Grammarly Premium to assist with language editing. To the best of our knowledge, we have no conflict of interest to declare.

Authors' contributions

T.H.: Conceptualisation, methodology, validation, writing – the initial draft. J.B.: Conceptualisation, writing – revisions, student supervision, project leadership. Both authors read and approved the final manuscript.

References

1. Statista. Socioeconomic indicators – South Africa [webpage on the Internet]. c2024 [cited 2023 Dec 11]. Available from: https://www.statista.com/outlook/co/socioeconomic-indicators/south-africa#:~:text=The%20total%20population%20share%20with,amount%20to%200.63%20in%202023
2. Sulla V, Zikhali P, Cuevas PF. Inequality in Southern Africa: An assessment of the Southern African Customs Union (English). Washington, DC: World Bank Group; 2023. https://doi.org/10.1596/37283
3. South African Government. Financial Sector Regulation Act 9 of 2017 [webpage on the Internet]. c2017 [cited 2023 Dec 11]. Available from: https://www.gov.za/documents/financial-sector-regulation-act-9-2017-english-sepedi-22-aug-2017-0000
4. World Bank. Universal financial access by 2020. c2018 [cited 2023 Dec 11]. Available from: https://www.worldbank.org/en/topic/financialinclusion/brief/achieving-universal-financial-access-by-2020
5. South African Government. Department of Home Affairs – Identity documents. No date [cited 2023 Dec 11]. Available from: https://www.dha.gov.za/index.php/civic-services/identity-documents
6. Rabaa'i AA, Zhu X, Jayaraman J. Understanding non-fungible tokens (NFTs): Overview, opportunities and challenges. In: Proceedings of the Northeast Business & Economics Association; Portsmouth, New Hampshire, USA. Worcester, MA: Northeast Business & Economics Association; 2022. p. 97–105. Available from: http://ez.sun.ac.za/login?url=https://www.proquest.com/scholarly-journals/understanding-non-fungible-tokens-nfts-overview/docview/2842095946/se-2.
7. Branson N, Garlick J, Lam D, Leibbrandt M. Education and inequality: The South African case (A Southern Africa Labour and Development Research Unit Working Paper Number 75). Cape Town: SALDRU, University of Cape Town; 2012.
8. Friderichs TJ, Keeton G, Rogan M. Decomposing the impact of human capital on household income inequality in South Africa: Is education a useful measure? Dev South Afr. 2023;40(5):997–1013. https://doi.org/10.1080/0376835X.2022.2163228
9. National Treasury. Full budget review 2022. Technical report, South African Government [document on the Internet]. c2022 [cited 2023 Dec 11]. Available from: https://www.treasury.gov.za/documents/national%20budget/2022/review/FullBR.pdf
10. South African Government. National Development Plan 2030: Executive summary [document on the Internet]. c2012 [cited 2023 Dec 11]. Available from: https://www.gov.za/sites/default/files/Executive%20Summary-NDP%202030%20-%20Our%20future%20-%20make%20it%20work.pdf
11. Ake C. The unique case of African democracy. Int Aff. 1993;69:239–244. https://doi.org/10.2307/2621592

12. Murisa T. Global philanthropy tracker: South Africa [document on the Internet]. c2020 [cited 2023 Dec 11]. Available from: https://scholarworks.iupui.edu/bitstream/handle/1805/26132/south-africa-report21.pdf

13. GPE Index. 2022 Global philanthropy environment index [webpage on the Internet]. No date [cited 2023 Dec 11]. Available from: https://globalindices.iupui.edu/environment-index/index.html

14. FinMark. FinScope SA pocket guide [webpage on the Internet]. c2019 [cited 2023 Dec 11]. Available from: https://finmark.org.za/system/documents/files/000/000/242/original/FinScope_SA_2019_Pocket_Guide_2020.pdf?1604679365

15. Demirgüç-Kunt A, Klapper L, Singer D, Ansar S, Singer D. The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19. Washington, DC: World Bank; 2022. http://hdl.handle.net/10986/37578

16. Better Than Cash Alliance. Define digital payments [webpage on the Internet]. No date [cited 2023 Dec 11]. Available from: https://www.betterthancash.org/define-digital-payments

17. BankservAfrica. Digital identity – A South African journey [webpage on the Internet]. c2021 [cited 2023 Dec 11]. Available from: https://www.bankservafrica.com/api/public/v2/filehandler/view/612f2d5e989a0e000128a550

18. Menotti EP, Farrell M. Vouchers: A hot ticket for reaching the poor and other special groups with voluntary family planning services. Glob Health Sci Pract. 2016;4:384–393. https://doi.org/10.9745/ghsp-d-16-00084

19. Valkama P, Bailey S. Vouchers as an alternative public sector funding system. Public Policy Admin. 2001;16(1):32–58. https://doi.org/10.1177/095207670101600103

20. Bohnenberger K. Money, vouchers, public infrastructures? A framework for sustainable welfare benefits. Sustainability. 2020;12(2), Art. #596. https://doi.org/10.3390/su12020596

21. World Bank. A guide to competitive vouchers in health. Washington, DC: The World Bank; 2004. Available from: http://documents.worldbank.org/curated/en/179801468324054125/A-guide-to-competitive-vouchers-in-health

22. Accenture. Meet me in the metaverse – Executive summary [document on the Internet]. c2022 [cited 2023 Dec 11]. Available from: https://www.accenture.com/content/dam/accenture/final/a-com-migration/thought-leadership-assets/accenture-meet-me-in-the-metaverse-executive-summary.pdf

23. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system [document on the Internet]. c2008 [cited 2023 Dec 11]. Available from: https://bitcoin.org/bitcoin.pdf

24. Sultan K, Ruhi U, Lakhani R. Conceptualizing blockchains: Characteristics & applications [preprint]. c2018 [cited 2023 Dec 11]. Available from: http://arxiv.org/abs/1806.03693

25. Bashir I. Mastering blockchain: A deep dive into distributed ledgers, consensus protocols, smart contracts, dApps, cryptocurrencies, Ethereum, and more. 3rd ed. Birmingham, UK: Packt Publishing; 2020.

26. Khan SN, Loukil F, Ghedira-Guegan C, Benkhelifa E, Bani-Hani A. Blockchain smart contracts: Applications, challenges, and future trends. Peer Peer Netw Appl. 2021;14:2901–2925. https://doi.org/10.1007/s12083-021-01127-0

27. Ante L. Smart contracts on the blockchain – A bibliometric analysis and review. Telemat Inform. 2021;57, Art.#101519. https://doi.org/10.1016/j.tele.2020.101519

28. Hsu C, Tu S, Huang Z. Design of an e-voucher system for supporting social welfare using blockchain technology. Sustainability. 2020;12(8), Art. #3362. https://doi.org/10.3390/su12083362

29. Kizza JM. Blockchains, cryptocurrency, and smart contracts technologies: Security considerations. In: Guide to computer network security. Texts in Computer Science. New York: Springer Cham; 2020. p. 575–600. https://doi.org/10.1007/978-3-031-47549-8

30. Lim SY, Fotsing P, Almasri A, Musa O, Mat K, Miss L, et al. Blockchain technology the identity management and authentication service disruptor: A survey. Int J Adv Sci Eng Inf Technol. 2018;8(4–2), Art. #1735. https://doi.org/10.18517/ijaseit.8.4-2.6838

31. Regner F, Schweizer A, Urbach N. NFTs in practice – Non-fungible tokens as core component of a blockchain-based event ticketing application. In: Krcmar H, Fedorowicz J, Boh W, Leimeister J, Wattal S, editors. Proceedings of the 40th International Conference on Information Systems (ICIS 2019); 2019 December 15–18; Munich, Germany. Atlanta, GA: Association for Information Systems; 2019.

32. Harraway T. Development of a peer-to-peer voucher donation management system enabled by blockchain [thesis]. Stellenbosch: Stellenbosch University; 2023.

AUTHORS:
Edwin E. Botha[1]
Kevin G. Harding[1]

AFFILIATION:
[1]Centre in Water Research and Development (CiWaRD), School of Chemical and Metallurgical Engineering, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Edwin Botha

EMAIL:
1600069@students.wits.ac.za

DATES:
Received: 26 June 2023
Revised: 24 May 2024
Accepted: 19 June 2024
Published: 04 Dec. 2024

HOW TO CITE:
Botha EE, Harding KG. Significance of international life cycle data in South African extended producer responsibility. S Afr J Sci. 2024;120(11/12), Art. #16384. https://doi.org/10.17159/sajs.2024/16384

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

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

EDITORS:
Michael Inggs
Thywill Dzogbewu

KEYWORDS:
extended producer responsibility, life cycle inventory, life cycle impact assessment, pulp and paper, South Africa

FUNDING:
Paper Manufacturers Association of South Africa

Significance of international life cycle data in South African extended producer responsibility

The South African extended producer's responsibility has made cradle-to-grave life cycle assessments a mandatory requirement for the paper and paper packaging industry. This is an intensive undertaking that requires a lot of data and time if primary data sets were to be created. The aim was to evaluate the applicability of using secondary and modified data sets in the life cycle assessment to speed up the process and reduce the amount of primary data required, with white-lined chipboard as the case study. Four white-lined chipboard data sets were used, a South African data set created from local industry data, a European data set from the Ecoinvent database and two modified European data sets, Scenario 1 and Scenario 2, to better represent the South African landscape. On an inventory level, the results indicated that the goal, scope and objective of the local and European life cycle assessments were similar, with minor differences. On an impact assessment level, the South African data had a much higher impact compared to the European data. This was mainly due to their reliance on fossil fuels for energy and electricity. On an uncertainty level, the uncertainty of the South African data was much higher, but this was due to the uncertainty related to the adjusted pedigree matrix and the cumulative nature of uncertainty in the life cycle inventory tiers. The results indicated that modified data sets with a base data set that has a similar goal and scope to the original South African study, and in which the data entries, data values and uncertainties are adjusted to match the South African process more closely would suffice.

Significance:

- International life cycle inventory data can be used to conduct local life cycle assessments provided that some minor modifications are made.
- In the South African context, the use of coal in the electric grid and boilers has the largest influence on the life cycle assessment outcomes.
- On a life cycle impact assessment (LCIA) level, uncertainty is cumulative, resulting in high uncertainty scores even if the variability of primary data (life cycle inventory level) is low.

Introduction

The South African government implemented the extended producer responsibility (EPR) legislation in 2021, with the intention of improving sustainability and the circular economy of the industries identified in the legislation. One of the ways in which they planned to achieve this is through life cycle assessment (LCA), with three objectives: reducing the volume of materials used, designing for recycling and reuse without compromising product functionality and reducing the environmental toxicity of the product in the final waste stream. The problem is that South Africa, unlike Europe and America, does not have sufficient life cycle inventories (LCIs) or databases, and, secondly, conducting LCAs with primary data collected from industries is both challenging and time consuming, especially considering the number of manufacturers and converters involved in the EPR's identified paper and paper packaging products (219 in the paper and paper packaging industry).

In this study, we aimed to conduct a representative LCA for white-lined chipboard (WLC) by using local, international and modified (both primary and secondary) data and adjusting the uncertainty of each data set to match its representativeness to the South African industry infrastructure. The objective was to compare the accuracy, representativeness and time it took to collect data and conduct the LCAs and to determine whether it is necessary to use 100% local data. The LCAs were conducted using SimaPro v9.3.0.3 simulation software and data from the Ecoinvent database, with the impacts determined using the ReCiPe 2016 v1.1 (H) impact method. The international ISO 14040 and ISO 14044 standards were used to guide the study.

Literature review

The pulp and paper industry is well established in LCA research. Although European countries and America have been using LCA for a long time, it is still an immature practice in South Africa (comparatively speaking). LCAs in South Africa span several industries, but to date, there are no LCA data sets available for the South African pulp, paper and paper packaging industries.[1,2] Therefore, South African researchers have three options available to conduct LCAs: option 1, to create the LCI themselves by approaching local companies; option 2, to use international LCI data; or option 3, to modify the international LCI data to create a mixture of local and international data. Each approach has its limitations and advantages.

Collecting data to create a locally representative LCI is time consuming and resource intensive.[3] The greater the scope of the LCA, the more detail and information will be required, and the longer the study will take.[4] Secondly, approaching local companies for data (especially if the LCA is not their intention) can prove to be a problem as they are reluctant to share sensitive information or refuse site visits, which could result in several data gaps and/or non-representative modelling of the process. However, despite these setbacks, using localised data is much more representative of the current technology, temporal boundary and geographical boundary and, if conducted accurately and correctly, will allow for LCI data that have much lower uncertainty and more reliability.[5]

International data sets are considered secondary because they have been created by previous researchers. By using secondary data, the time spent on creating the LCI is significantly reduced as the data can be extracted from a database such as Ecoinvent and the US Life Cycle Inventory (USLCI). This could allow for a study with a greater scope and cover a larger value chain in a shorter time. Also, using secondary data could give a researcher a generalised idea of the overall impact of a process/product and can help identify potential hotspots. This could guide the goal and scope of a future study, which would then focus on obtaining primary data for those specific hotspots or observed anomalies.[4] Several drawbacks also accompany the use of secondary data[3], such as data gaps, being unrepresentative on a local level and may not conform to the scope of the local study. These factors not only influence the outcome of the LCA but also increase the level of uncertainty of the LCI[3,6,7], which reduces the confidence in the data and the LCA results.

The third option is creating a modified data set. These data sets consist of a mixture of primary and secondary data. It is not possible to conduct an LCA using LCA software without the use of secondary data. For example, when conducting an LCA using the Ecoinvent database, it should be noted that even a South African data set, such as pulp manufacturing, is built with other data sets, such as transport, electricity, machinery, fuels, and chemicals. Although Ecoinvent does contain South African data sets for electricity and transport, it does not contain South African data sets for the chemicals used in pulping, such as sodium hydroxide solution (white liquor), so international data sets are used to represent those chemicals. The only changes that can then be made are to the data values and uncertainty.

Based on this, there are three ways of creating a modified data set. One is to switch a process/product/material entry to its local counterpart. The second is to replace the non-local data value (secondary data) with a primary data value representing the local process. The third then is a mixture of both. Creating modified data sets is faster than creating primary data sets, they are more representative than the secondary data option, and sensitive data are indirectly accounted for as the values in the base data set are averaged for that country or region, which could help bridge local data gaps. The downside of using modified data sets is similar to the secondary data set and revolves around the lack of local representativeness, increased uncertainty, lack of reliability, the potential of over/underestimating the LCA results, and lastly, differences in the goal and scope.[8,9]

The aim of this study was therefore to test the applicability and representatives of the three modelling techniques, to indicate which one is more favoured on a time scale and an LCI, life cycle impact assessment (LCIA) and data quality level.

Methods

Data set preparation and alterations

Four data sets were created and evaluated, namely:

1. South African
2. European
3. Scenario 1
4. Scenario 2

South African data for eucalyptus forestry, kraft pulping and WLC manufacturing were collected. The data were collected from local companies and consisted of both primary and generalised South African data. The primary data covered a 3-year temporal boundary from 2018 to 2020.

The European WLC data set, 'White Lined Chipboard Carton Production, RER, 2018–2018', was extracted from the Ecoinvent database and was used as the secondary data set. No changes were made to this data set. The European data set represents 52% of all European mills. Some of the mills were integrated, performing pulping, manufacturing and converting on the same site, and some of them were self-standing manufacturing mills. This means that the European WLC data set includes some data entries only used in pulping (wood chips, pulpwood, sodium hydroxide, sodium sulfate, calcium carbonate, peroxide, etc.); however, this was accounted for by including South African Kraft pulping.

The modified data sets were derived from the European data set. For Scenario 1, all the European data entries with relevant South African counterparts were replaced, but no changes were made to the numerical entries. The data entries changed included coal, electricity, water, effluent, natural gas, transport, diesel, and light fuel oil. The distribution of the data (uncertainty) was also changed based on its representativeness of the South African process via the Ecoinvent pedigree matrix. The matrix was used to adjust the standard deviation by grading the representativeness of the data on a scale of 1 to 5 by considering the data's reliability, completeness, temporal correlations, geographical correlations and technological correlations.

In Scenario 2, the same process as Scenario 1 was followed; however, the numerical values of the replaced data entries (and others) were also substituted with South African primary data and their uncertainty was modelled using the standard deviation of the South African data. These changes were made to virgin fibre, recycled fibre, coal, electricity, water, effluent, natural gas, transport, diesel, light fuel oil, air emissions (CO_2, SO_2, NO_x, CO, PM) and effluent (chemical oxygen demand, suspended solids, biological oxygen demand, total organic carbon) composition, ash, sludge, and landfill waste. Scenario 2 was modified to convert data entries to their respective South African counterparts and to model data that mills are most likely to supply (non-sensitive data). This allowed the European data to account for sensitive data and data gaps, including bills of materials, process chemicals, emissions other than those listed above, vehicle maintenance and process maintenance. Data sets that were not in any way applicable to South Africa were removed. Changes made to the EU data set for Scenario 1 and Scenario 2 can be found in the supplementary material.

Modelling of scenarios

The comparisons for the data sets were conducted using the Ecoinvent allocation cut-off by classification technique. This technique essentially corresponds to the 'polluter pays' approach, thus incentivising the use of recyclable products and increased recycling. This is in line with the EPR, which expects manufacturers to take responsibility for their products in the final waste stream and reduce the number of raw materials used in their products. The LCA was conducted using SimaPro v 9.3.0.3 (2022) and data from EcoInvent[10], with the impact assessment calculated using ReCiPe 2016 v1.1 Midpoint (H).

The goal for the South African WLC LCA

The production of WLC (unconverted) was studied using a cradle-to-gate assessment and included eucalyptus forestry, kraft pulping and WLC manufacturing. The mass and energy flows were used to determine the environmental impact of producing 1 kg of WLC (unconverted). This value was used as the 'functional unit' of the study to which all other flows and co-products were scaled. The comparisons took place in three steps. The first comparison was an inventory comparison between the South African and European data sets. The second comparison was an LCIA comparison for the South African, European, and Scenario 1 and Scenario 2 modified data sets. The last comparison was a Monte Carlo analysis for all four data sets.

The scope for the South African WLC LCA

Forestry

The forestry process was for pulp logs only. The forestry portion of the value chain starts with the planting of eucalyptus seedlings and ends with the logs being transported to the pulping mill (Table 1). Forestry was modelled using economic allocation, and all the weight of forestry was allocated to the debarked logs. The seed germination and nursery process was excluded from the study. For forestry, processes such as pre-planting fertilisation, pre-planting pesticide and herbicide application, pre-planting irrigation, and the continuation of these throughout the rotation period were excluded as the application of these is often small, inconsistent and conditional, making it difficult to track and numerically represent. The transport from the point of harvest to the roadside pick-up was considered negligible.

Table 1: South African eucalyptus forestry process flows for 1 m^3 pulp logs harvested

Product	Value	Unit	Source	Note
Wood under bark	1	m^3		Reference product
Resources				
Carbon dioxide	898	kg	Secondary	Absorbed from environment
Energy (biomass)	12 939	MJ	Secondary	
Wood	1	m^3	Primary	Harvested
Forest land occupied	567	m^2a	Primary	Land area used for a period of tree growth
Rail/road embankment land occupied	17.0	m^2a	Secondary	Land used for rail and road during the period of tree growth
Land transformed into forestry	0.21	m^2	Secondary	Afforestation
Land transformed from forestry	2.5	m^2	Secondary	Deforestation
Land transformed from traffic area, rail/road embankment	0.075	m^2	Secondary	Roads removed during deforestation
Land transformed into a traffic area, rail/road embankment	0.0063	m^2	Secondary	Roads built during afforestation
Materials/fuels				
Harvesting, forestry harvester	0.03	h	Secondary	Machinery used to harvest trees
Forwarding, forwarder	0.038	h	Secondary	
Skidding, skidder	0.077	h	Secondary	
Power sawing, without a catalytic converter	0.25	h	Secondary	
Gravel, crushed	20.0	kg	Secondary	Road construction and maintenance
Lubricating oil	0.3	kg	Secondary	Maintenance
Diesel	78	MJ	Secondary	Fuels used in machinery

Pulping

The pulping portion of the value chain starts with the receipt of debarked logs at the pulping mill and ends with pulp bales transported to the manufacturing mill. The process included wood cutting and chipping, pulping, drying, bleaching, baling, transport, effluent treatment and disposal, steam and electricity generation (from coal, wood yard waste, and chemical recovery), chemical recovery and waste treatment (Table 2). Mass allocation was used. The impact of capital equipment, buildings and their maintenance was considered negligible. On-site transport was not considered. Business travel and employee commutes were considered negligible.

Manufacturing

The manufacturing portion of the value chain starts with the receipt of eucalyptus virgin pulp bales and recycled paper grades and ends with the production of WLC jumbo reels (unconverted). The process of WLC manufacturing includes re-pulping of virgin and recycled pulp, screening and treatment, calendaring, coating, on-site energy generation, drying, water treatment, on-site transport, transport to clients, and cutting and wrapping of WLC jumbo reels. The LCI for WLC is presented in Table 3. Mass allocation was used, and the impact of WLC manufacturing was solely attributed to WLC (no co-products). The same assumptions as for the pulping section were used.

Results

Inventory comparison

In the analysis, the inventory data entries were compared by the number of data entries in both the South African and European data sets and by studying the impact of the processes when energy and electricity data entries were excluded (Figure 1), primarily due to the country-specific influence of fuel sources.

The information in Table 4 indicates a minor difference between the two data sets when considering the number of data entries. Several mills were included because the European data set represented 52% of Europe's WLC production. Because each mill uses different chemicals and raw material sources, there are some flows represented by two or more data entries. For example, the European mill had three water entries (tap, well and lake) compared to the one (tap) in the South African mill. Similarly, because some mills were integrated, the additional pulping chemicals and waste are also part of the LCI (but negligible in mass at <1% in total). Hence, there are differences in the number of data entries. With this considered, the data coverage for the South African LCI is good compared to the European LCI.

The significant mass balance difference between the two data sets exists because the European mills use and discharge more water. The European mill uses around 18 L of fresh water per kg WLC produced and discharges 15.9 L per kg WLC, whereas the South African mill uses 8.8 L of fresh water per kg WLC and discharges 5.7 L per kg WLC. As the water volumes are scaled to the reference product, the difference in consumption is quite notable. However, their use of water is reasonable, because their process includes pulping, manufacturing and converting; the location of mills in the European countries is in water-abundant regions, unlike South Africa, which is a semi-arid, water-scarce country where water recycling is crucial (up to 80%). Secondly, as the European mills are integrated, there is an additional mass of wood chips, pulping chemicals, pulping wastes (white liquor, bleaching, dreg wastes, lime mud, etc.), inks and converting wastes. However, the greatest reason for the mass balance difference was the volume of fresh water used and effluent discharged.

Two approaches were taken to study the impact of the LCI. In the first approach, the boiler and electricity data sets were excluded from the analysis to solely study the impact of material use, on-site waste

Table 2: South African eucalyptus kraft pulping process flows for 1 kg baled pulp produced

Product/process	Value	Standard deviation	Unit	Source	Notes
Materials/fuels					
Delignification and bleaching	0.011	0.0039	kg	Primary	Chemicals
pH control	0.008	0.0002	kg	Primary	Chemicals
Heavy fuel oil	0.14	0.071	MJ	Primary	Used for drying and electricity production
Heat (biomass)	0.77	0.028	MJ	Primary	Used for drying and electricity production
Disinfectant	9.3×10^{-5}	2.1×10^{-5}	kg	Primary	Chemicals
Starch	0.0019	6.8×10^{-4}	kg	Primary	Chemicals
Other fillers and additives	0.019	0.0057	kg	Primary	Chemicals
Sizing agent	0.00014	3.0×10^{-5}	kg	Primary	Chemicals
Pitch control	0.00022	8.4×10^{-5}	kg	Primary	Chemicals
Eucalyptus (wood under bark)	0.0031	–	m^3	Primary	From forests
Water	25	1.3	kg	Primary	Surface water source
Electricity/heat					
Electricity, natural gas	0.11	0.008	kWh	Primary	On-site
Electricity, wood bark	0.039	0.0028	kWh	Primary	On-site
Electricity, coal	0.19	0.014	kWh	Primary	On-site
Electricity, coal	0.034	0.0053	kWh	Primary	Municipality
Heat, black liquor	6.1	0.32	MJ	Primary	On-site
Transport from forests (40- ton truck)	100	–	kgkm	Primary	
Emissions to water					
Chemical oxygen demand	0.00068	2.9×10^{-4}	kg	Primary	To the ocean, after treatment
Suspended solids	0.00020	8.6×10^{-5}	kg	Primary	
Nitrogen	7.9×10^{-7}	2.8×10^{-7}	kg	Primary	
Phosphorus	1.1×10^{-6}	7.3×10^{-7}	kg	Primary	
Waste to treatment					
Wood ash mixture	0.00011	5.3×10^{-5}	kg	Primary	Landfill
Hard coal ash	0.00055	2.6×10^{-4}	kg	Primary	Landfill
Municipal solid waste	0.001	3.7×10^{-4}	kg	Primary	Landfill
Hazardous waste	2.3×10^{-5}	2.1×10^{-5}	kg	Primary	Incineration
Green liquor dregs	0.00077	2.2×10^{-4}	kg	Primary	Landfill
Sludge	0.004	–	kg	Primary	Landfill

generation and transport. The boiler and electricity data sets were excluded because both mills used different fuel sources, the electric grids for both regions were considerably different, and they were the two processes with the most significant influence.

Comparing the LCI for the two processes, excluding electricity and boiler data sets, the results showed that the impact on the rest of the inventory was quite similar (Figure 1). The European impact was slightly higher due to additional data entries; nevertheless, the results from Figure 1 show that if a secondary data source was well modelled and included a similar goal and scope, the impact of process-specific data would be similar. This is especially true for the pulp and paper industry, where many processes follow the same pulping and manufacturing processes and use similar chemicals, even though they are from different regions. The only significant differences were for freshwater ecotoxicity and eutrophication, but this was because the European mills discharged their effluent into surface water sources, their water balances were higher (larger mass of contaminants), and they reported more contaminants.

In the second scenario, the inclusion of the boiler and electricity processes completely changed the results (Figure 2). The environmental impact of using coal in the South African scenario has caused the environmental impact of the South African data set to increase and surpass that of the European mill's impact for all impact categories. The electricity generation process is country-specific and can vary extensively. In Europe, 15% of fossil fuels are used to supply electricity to the grid, whereas in South Africa, 85% of fossil fuels are used to supply their electric grid. The same applies to boilers because the boiler's fuel source depends on the most

Table 3: South African white-lined chipboard (WLC) production process flows for 1 kg WLC produced

Product	Value	Standard deviation	Unit
Raw materials			
Virgin kraft fibre	0.19	0.0052	kg
Virgin mechanical fibre	0.06	0.0017	kg
Recycled fibre	0.85	0.0096	kg
Fresh water	8.8	0.27	L
Energy			
Electricity	1.1	0.065	kWh
Natural gas	0.76	0.12	MJ
Diesel	0.13	–	MJ
Coal	0.36	0.026	kg
Chemicals			
Polymers	0.25	0.013	kg
Salts	0.048	0.0013	kg
Starches	0.018	0.00028	kg
Amines	0.00013	1.2×10^{-5}	kg
Sodium hydroxide	0.0011	3.3×10^{-5}	kg
Defoamers	2.8×10^{-4}	7.8×10^{-6}	kg
Sizing agents	0.0004	0.00016	kg
Minerals	0.026	0.002	kg
Optical brighteners	0.00015	4.2×10^{-6}	kg
Biocide	0.0011	2.6×10^{-5}	kg
Water treatment	0.00023	3.0×10^{-4}	kg
Boiler water softening	4.0×10^{-4}	5.4×10^{-5}	kg
Transport			
By ship	849	24	kg.km
Transport to mill	290	13	kg.km
Transport from mill	420	40	kg.km
Solid wastes			
Screened waste	0.076	0.016	kg
Effluent	5.7	0.41	L
Sludge	0.048	0.0083	kg
Ash	0.085	0.018	kg
Emissions			
Water lost as steam during drying	1.5	–	L
Carbon dioxide, fossil	0.6	0.021	kg
Carbon monoxide, fossil	0.0015	5.7×10^{-4}	kg
Methane, fossil	1.0×10^{-4}	–	kg
Nitrogen oxides	0.0012	1.7×10^{-4}	kg
Particulate matter	0.001	1.0×10^{-4}	kg
NMVOC	1.8×10^{-5}	–	kg
Sulfur dioxide	0.0019	0.00032	kg
Metals	0.0027	–	kg
Inorganics	4.2×10^{-4}	–	kg
Organics	1.7×10^{-4}	–	kg

Table 4: Differences between the South African and European white-lined chipboard data entries

Data entries	Unit	South African	European
Inputs	–	32	40
Outputs (including emissions)	–	83	98
Outputs (excluding emissions)	–	13	18
Chemical inputs (potentially sensitive data)	–	20	18
Recycled fibre	–	3	2
Virgin fibre	–	2	3
Energy sources (all)	–	3	5
Energy sources (fossil fuel)	–	3	3
Mass in	kg	10.69	19.5
Mass out	kg	10.65	20.22
Mass in (excluding water)	kg	1.89	1.5
Mass out (excluding effluent)	kg	4.95	4.32

abundant fuel in that region or country. The integrated European WLC mills used a mix of coal and biomass, but biomass was the most used fuel, whereas the South African WLC mill used 100% coal because it was the closest, most abundant and most affordable fuel source. The results indicate that when using secondary data, it is important to consider country- or region-specific processes and processes that result in significant changes from one region or country to another.

Life cycle impact assessment

The results of the LCIA for the South African WLC, European WLC, and Scenarios 1 and 2 are presented in Figure 3. Scenario 1 was modelled by changing data entries in the European data set with available South African SimaPro data entries and adjusting the uncertainty of the data entries. Scenario 2 was modelled the same as Scenario 1, but the data values were also changed based on the likelihood that a company would be willing to supply that kind of information (non-sensitive data).

The results for Scenario 2 indicated that the modified data set could be used to represent the South African process. Even though Scenario 2 had the closest representation of the South African WLC manufacturing process, it still overestimated the impact in all categories, particularly for eutrophication and freshwater ecotoxicity. This overestimation was anticipated as the base data set was European, and, as observed in Figures 1 and 2, the impact of eutrophication and freshwater ecotoxicity was fairly high for the European data set, even when the electricity and boiler data were included.

Apart from the eutrophication impact outlier, Scenario 2 could be used instead of creating an entirely new primary LCI. However, Scenario 2 should be applied cautiously, as the goal, scope and inventory completeness of the base data sets should be the same or very similar to the goal and scope of the South African LCA; otherwise, the results from the LCIA would be non-representative and futile.

Regarding Scenario 1, it was beginning to show a closer resemblance to the South African data set. Even though the differences were still large, it does indicate that even small changes, such as converting data entries to their respective South African equivalents, can already start shaping the data set and the LCIA to be more locally representative.

The results for the European data set indicate that using an international data set without some way of linking it or adjusting it to match South African processes, as with Scenario 2, is not acceptable. There is a large difference between the European and South African data sets,

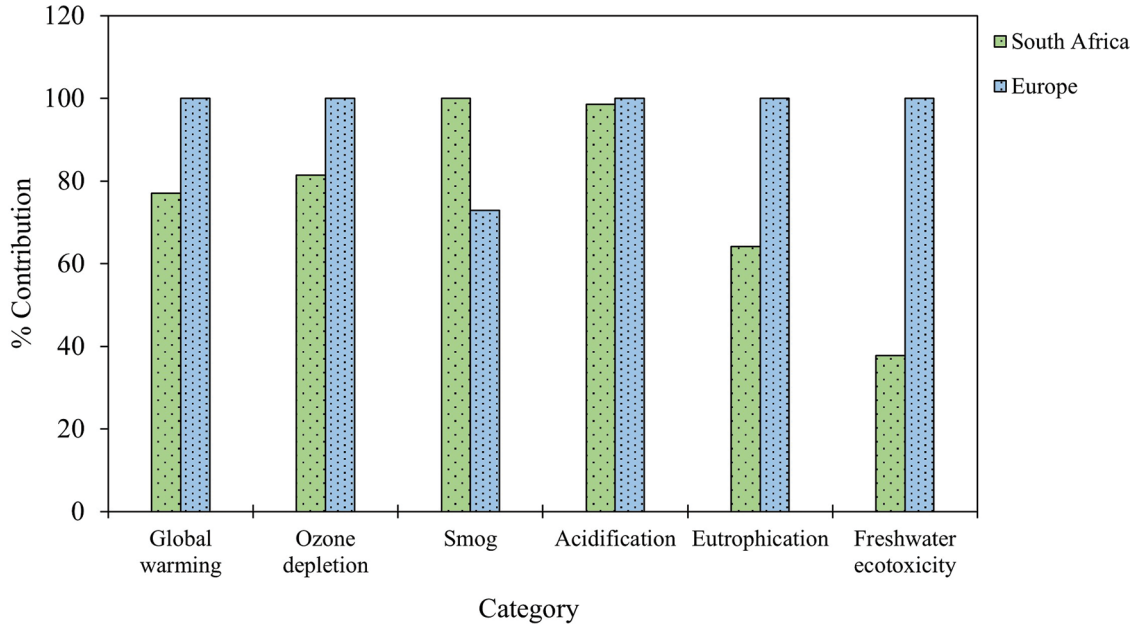


Figure 1: Life cycle impact assessment of white-lined chipboard manufacturing with the electricity and boiler data sets excluded, indicating the influence of process-specific data.

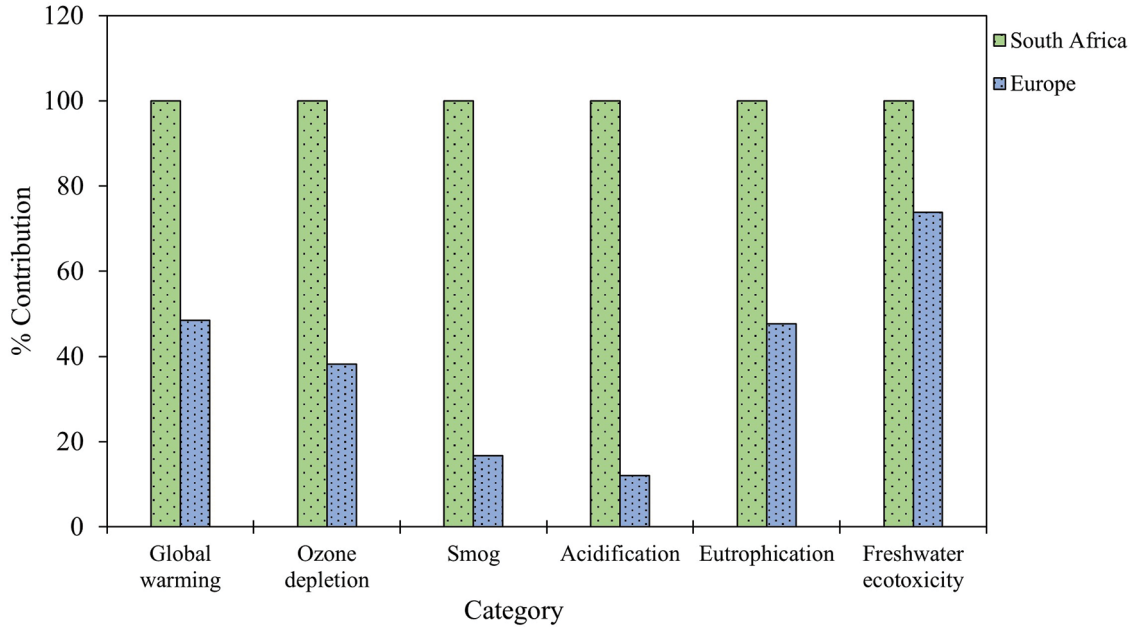


Figure 2: Life cycle impact assessment of white-lined chipboard manufacturing with the electricity and boiler data sets included indicating the influence of country/region-specific data.

which shows that using international data to represent local processes could lead to large over- or underestimations that are inaccurate and not representative of the local process.

Uncertainty tests

In building the LCI for a study, two aspects of data collection need to be considered: the data type and data quality. Data type was covered in the previous two sections by observing the influence of primary data, secondary and hybrid (mixed) data on the LCIA results. Data quality refers to the data uncertainty (representativeness), variability and model assumptions.[11]

Data variability was addressed in SimaPro using standard deviation and the Ecoinvent pedigree matrix. The results from the Monte Carlo analysis for each of the data sets are presented in Figure 4, showing the impact of uncertainty on the outcome of the data and the level of variability.

Considering the South African data set, two observations relating to the uncertainty of the data sets could be made. Firstly, the uncertainty for the South African data set was higher in comparison to the European data set. Secondly, the uncertainty for the two water impact categories, eutrophication and freshwater ecotoxicity, was considerably higher compared to the other impact categories.

Concerning the first observation. One of the problems with the South African process is the lack of data sets in Ecoinvent (and the lack of South African data sets in general). SimaPro data sets are then built using non-South African data entries. This increases the uncertainty as the representativeness of the data set is lower, which increases the pedigree matrix score, which in turn increases the standard deviation value. Even though the standard deviation (variability) of the collected primary data in the LCI was low, the uncertainty is higher via the simulation results.

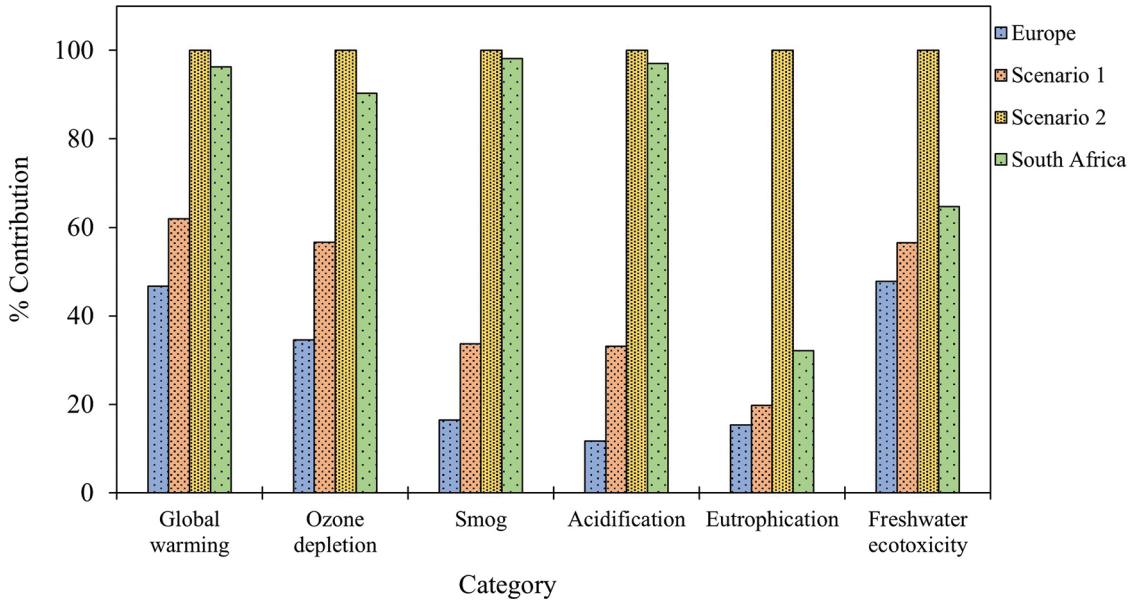


Figure 3: Life cycle impact assessment results for the comparison of South African white-lined chipboard, European white-lined chipboard and the two modified data sets, Scenario 1 and Scenario 2.

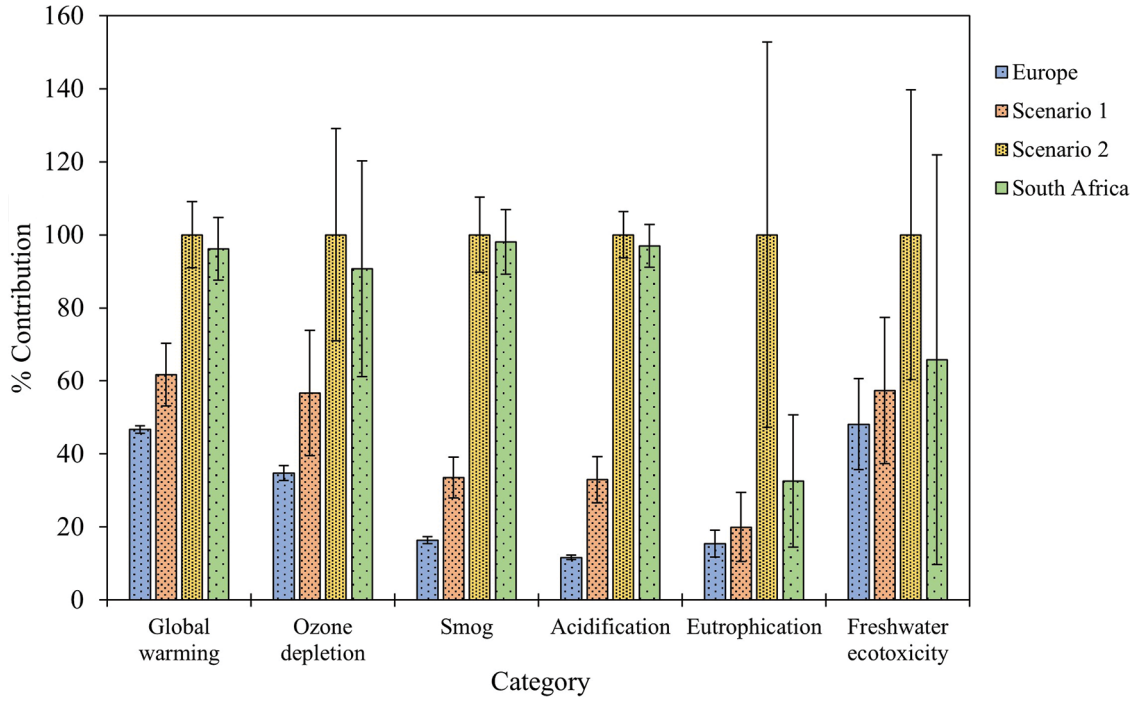


Figure 4: Uncertainty of environmental impacts scaled to one, with the error bars representing the standard deviation.

Additionally, the data set in SimaPro consists of tiers (process tree). For example, the WLC manufacturing data set consists of data sets (second tier), and those data sets are built using more data sets (third tier) and so on. The final uncertainty generated by the Monte Carlo analysis is not the uncertainty of the WLC data set itself, but a cumulative value of all the uncertainties across all the tiers.[12] This means that a South African data entry, such as electricity, becomes a second-tier data entry in the WLC data set. This data entry is also built with South African and non-South African data, which means the standard deviation will be higher for the non-South African data values. This then cumulatively adds to the WLC data set's uncertainty as well. Hence, the increased uncertainty via the pedigree matrix for non-South African data entries and the tiered nature of the data sets results in the higher uncertainty observed in Figure 4.

As for the second observation, the uncertainty for all the data sets was highest for freshwater ecotoxicity and highest for the South African and Scenario 2 eutrophication relative to the other impact categories, especially considering the significant differences among the different data sets. A study by Chen et al.[13] indicated that uncertainties can be caused by the LCI, substance coverage by the different impact methods (impact method inventory) and differences in the characterisation factor values of the substances. Therefore, the reason for this observation could be uncertainty in the inventory for the ReCiPe 2016 impact method. This indicates that the uncertainty in an LCA is linked to both the uncertainty in the LCI and the uncertainty in the impact method. Similar conclusions were made by Alyaseri and Zhou[14] and by Barahmand and Eikeland[7]. Therefore, the greater uncertainty for these two impact categories could be associated with the inventory of the ReCiPe 2016 impact method and not the inventory or the uncertainty of the LCIs.

However, the European data set had the smallest overall uncertainty associated with the impact values, as the tiered data sets used were

European data sets, which meant that its representativeness was good, and the associated uncertainty and the cumulative uncertainty values were low. For Scenario 1, a similar observation could be made, where the overall uncertainty remained small. In Scenario 2, the uncertainty became more like the South African data set because the data set entries were changed, the pedigree matrix was updated, and the data values were also changed, allowing Scenario 2 to become more South African, hence the larger impacts and greater uncertainty.

Concerning the application of uncertainty in the EPR, if the variability of the primary data in the LCI is low and the information gathered is representative of the local geography, infrastructure, technology and temporal correlation, then uncertainty should not be an influence and only considered on an LCI level.

Conclusion

Based on the EPR objectives, international LCI data can be used to conduct local LCAs for the paper industry, provided that the base data set used has a similar goal, scope and modelling choices and that, as a minimum, the electricity, boiler fuel and emissions are replaced by its South African data entry counterpart and primary data.

The tiered nature of LCA causes uncertainty to be cumulative and therefore much more significant on the LCIA level; however, if the variability of primary data used is low, uncertainty should be considered on an LCI level and not the LCIA level.

Generally, it is assumed that the processes of the paper and paper packaging industry have little variation from one country to another, and as a result, these findings could potentially apply to other paper and paper packaging products defined by the EPR. This will only be possible if the base data set is appropriately modified and has the same or similar goal and scope as its South African counterpart.

Acknowledgements

We thank the Paper Manufacturers Association of South Africa (PAMSA) and their partner companies for providing the data used in the LCA.

Funding

This research was funded by the Paper Manufacturers Association of South Africa (PAMSA).

Data availability

All the data supporting the results of this study are included in the article itself and the supplementary material. The data tables presented in the main text are the South African inventory for forest, pulping and manufacturing. The supplementary material includes the modifications to the RER WLC data set for Scenario 1 and Scenario 2, which includes changes in uncertainty and data values.

Declarations

We have no competing interests to declare. We have no AI or LLM use to declare.

Authors' contributions

E.E.B.: Conceptualisation, data collection, data analysis, writing – the initial draft. K.G.H.: Student supervision, writing revisions. Both authors read and approved the final manuscript.

References

1. Harding KG, Friedrich E, Jordaan H, le Roux B, Notten P, Russo V, et al. Status and prospects of life cycle assessments and carbon and water footprinting studies in South Africa. Int J Life Cycle Assess. 2021;26(1):26–49. https://doi.org/10.1007/s11367-020-01839-0
2. Karkour S, Rachid S, Maaoui M, Lin CC, Itsubo N. Status of life cycle assessment (LCA) in Africa. Environments. 2021;8(2):10–56. https://doi.org/10.3390/environments8020010
3. Ciroth A, Di Noi C, Burhan SS, Srocka M. LCA database creation: Current challenges and the way forward. Int J Life Cycle Assess. 2019;3(2):41–51. https://doi.org/10.52394/ijolcas.v3i2.105
4. Quist Z. Primary vs. secondary data – What's the best in LCA? [webpage on the Internet]. c2022 [cited 2023 Jan 25]. Available from: https://ecochain.com/knowledge/primary-vs-secondary-data-in-lca/
5. Kalverkamp M, Helmers E, Pehlken A. Impacts of life cycle inventory databases on life cycle assessments: A review by means of a drivetrain case study. J Clean Prod. 2020;269, Art. #121329. https://doi.org/10.1016/j.jclepro.2020.121329
6. Bamber N, Turner I, Arulnathan V, Li Y, Zargar Ershadi S, Smart A, et al. Comparing sources and analysis of uncertainty in consequential and attributional life cycle assessment: Review of current practice and recommendations. Int J Life Cycle Assess. 2020;25:168–180. https://doi.org/10.1007/s11367-019-01663-1
7. Barahmand Z, Eikeland MS. Life cycle assessment under uncertainty: A scoping review. World. 2022;3:692–717. https://doi.org/10.3390/world3030039
8. Canals LMI, Azapagic A, Doka G, Jefferies D, King H, Mutel C, et al. Approaches for addressing life cycle assessment data gaps for bio-based products. J Ind Ecol. 2011;15(5):707–725. https://doi.org/10.1111/j.1530-9290.2011.00369.x
9. Ossés de Eicker M, Hischier R, Kulay LA, Lehmann M, Zah R, Hurni H. The applicability of non-local LCI data for LCA. Environ Impact Assess Rev. 2010;30(3):192–199. https://doi.org/10.1016/j.eiar.2009.08.007
10. Wernet G, Bauer C, Steubing B, Reinhard J, Moreno-Ruiz E, Weidema B. The ecoinvent database version 3 (part I): Overview and methodology. Int J Life Cycle Assess. 2016;21(9):1218–1230. https://doi.org/10.1007/s11367-016-1087-8
11. Soares SR, Finotti AR, Prudêncio da Silva V, Alvarenga RAF. Applications of life cycle assessment and cost analysis in health care waste management. Waste Manag. 2013;33(1):175–183. https://doi.org/10.1016/j.wasman.2012.09.021
12. Goedkoop M, Oele M, Leijting L, Ponsioen T, Meijer E. Introduction to LCA with SimaPro [webpage on the Internet]. c2016 [cited 2023 Jan 25]. p. 69. Available from: https://support.simapro.com/s/article/Introduction-to-LCA
13. Chen X, Matthews HS, Griffin WM. Uncertainty caused by life cycle impact assessment methods: Case studies in process-based LCI databases. Resour Conserv Recycl. 2021;172, Art. #105678. https://doi.org/10.1016/j.resconrec.2021.105678
14. Alyaseri I, Zhou J. Handling uncertainties inherited in life cycle inventory and life cycle impact assessment method for improved life cycle assessment of wastewater sludge treatment. Heliyon. 2019;5(11), e02793. https://doi.org/10.1016/j.heliyon.2019.e02793

Ambient seismic noise imaging of a tailings dam internal structure

AUTHORS:
Thulisile Kunjwa[1,2]
Moctar Doucouré[1]

AFFILIATIONS:
[1]Africa Earth Observatory Network (AEON), Nelson Mandela University, Gqeberha, South Africa
[2]Department of Geosciences, Nelson Mandela University, Gqeberha, South Africa

CORRESPONDENCE TO:
Thulisile Kunjwa

EMAIL:
kunjwat@gmail.com

DATES:
Received: 23 June 2023
Revised: 27 Mar. 2024
Accepted: 19 June 2024
Published: 04 Dec. 2024

HOW TO CITE:
Kunjwa T, Doucouré M. Ambient seismic noise imaging of a tailings dam internal structure. S Afr J Sci. 2024;120(11/12), Art. #16366. https://doi.org/10.17159/sajs.2024/16366

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITORS:
Michael Inggs
Thywill Dzogbewu

KEYWORDS:
tailings, seismic interferometry, ambient seismic noise, Green's function, imaging

FUNDING:
South African National Research Foundation

Several tailings dam failures have been reported over the past few decades, raising questions about the stability of these structures. Over the past 30 years, there have been more severe failures of tailings dams, which have resulted in damage to the environment, fatalities and severe socio-economic issues. It is often impossible to predict when tailings dams could fail, so there is an urgent need to develop cost-effective methods for monitoring their stability and preventing such catastrophic events. Recent advancements in techniques related to ambient seismic noise have the potential to introduce new approaches to subsurface imaging and monitoring. In this study, we investigated the potential of using ambient seismic noise to monitor the interior wall of a tailings dam. We used 20 3C short-period geophones to record ambient noise over 3 days continuously. We set up the geophones at the Harmony Gold Mine tailings dam in Welkom, South Africa, along a survey profile approximately 100 m long. Seismic interferometry was used from the recorded data to retrieve empirical Green's functions. We computed the dispersion curves' inversion to determine the dam wall's shear wave velocity at different depths. The calculated shear wave velocity cross-sections revealed a region of reduced velocity within the dam wall, situated between 2 m and 10 m beneath the surface. This zone of low velocity can be a sign of water-saturated material or other subsurface anomalies, which might jeopardise the dam's structure.

Significance:

Tailings dams have a long history of ruptures and collapse. The results from this study indicate that passive seismic interferometry can be a cost-effective tool for imaging and identifying minor changes within the interior of the tailings dam wall. This technique could prove valuable in detecting potential instabilities before they reach a critical stage. This can allow for early intervention and remedial measures, such as reinforcing the dam wall or adjusting the water levels in the tailings dam.

Introduction

Tailings dams store waste after separating ore and gangue in mining operations. Some dams reach dozens of metres in height and stretch over several kilometres.[1] These structures can fail, resulting in catastrophic environmental and social consequences. When a tailings dam collapses, it can release millions of tons of toxic waste into surrounding ecosystems, contaminating water sources, and damaging infrastructure. Estimating the exact number of tailings dams worldwide is challenging, as many operate without regulation or documentation. An estimated 3500 large dams exist worldwide; a significant portion are dedicated to supporting mining operations and can potentially be classified as tailings dams.[2]

South Africa has the highest number of environmentally hazardous tailings dams globally due to the predominance of upstream tailings dams, which are considered unsafe and are banned in some countries.[3] There are over 212 reported tailings dams in South Africa[4], and there have been serious dam failures recorded over the past three decades.

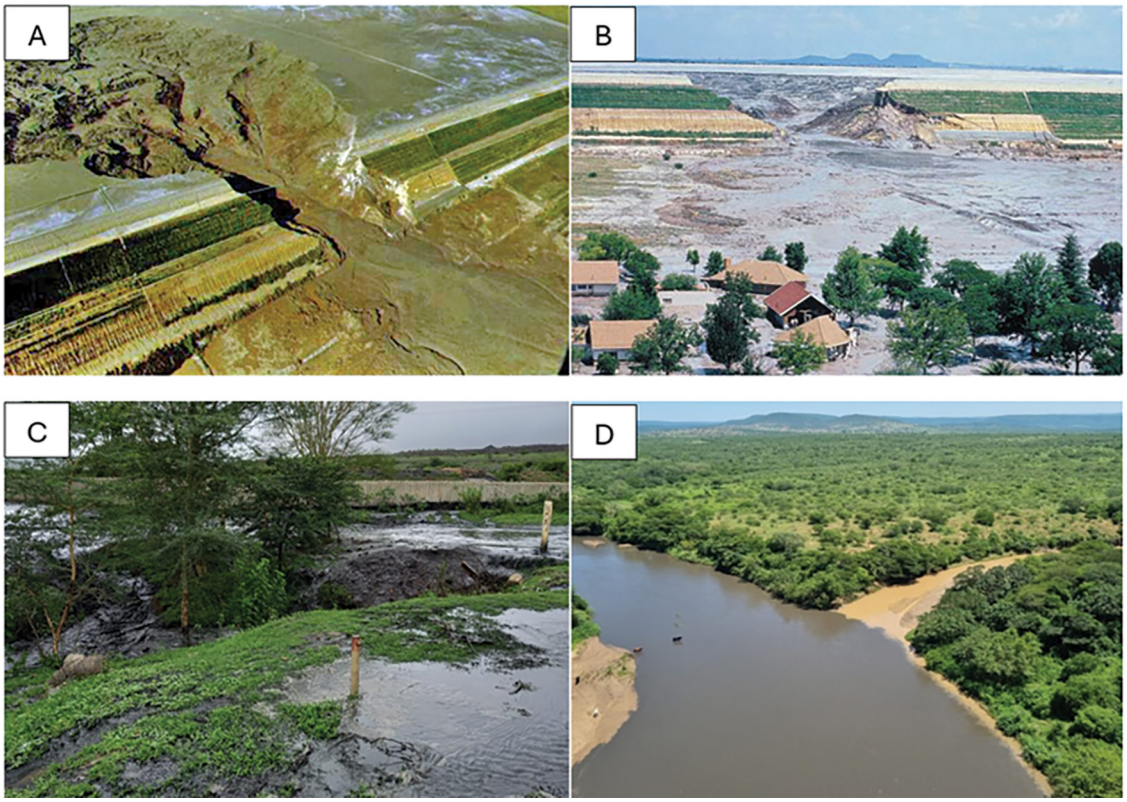
On 22 February 1994, the Merriespruit tailings dam collapsed, resulting in the discharge of 600 000 m^3 of tailings and 90 000 m^3 of tailings water that spilled into the town of Merriespruit. This incident is considered as the most infamous tailings dam accident in South Africa. After a few hours of heavy rain, the dam overtopped and broke, resulting in static liquefaction and the flow slide of a section of the embankment as soon as sufficient toe material had eroded. The disaster resulted in the death of 17 people and the destruction of thousands of houses. Figure 1A and 1B shows an aerial view of the dam after it collapsed and images of the town to which the tailings flowed.

On 24 December 2021, coal mining effluent poured from a collapsed dam at the Zululand Anthracite Colliery (ZAC) mine, causing another tailings dam to break. At least 1 500 000 L of contaminated mining waste flowed close to residences before reaching the Black Umfolozi River, which flows through subsistence farming areas and the Hluhluwe-iMfolozi Game Reserve, where the community depends on the river for drinking water and for watering their crops. Figure 1C and 1D shows areas contaminated by the toxic and acidic coal mine waste released into the rivers that flowed through rural villages, the Hluhluwe-iMfolozi and iSimangaliso nature reserves, and other nearby locations.

The most recent dam failure happened at Jagersfontein in South Africa on 11 September 2022 – most of the tailings discharged from the south of the dam, where there was a breach. The width of the plume reached up to 1.5 km and extended about 8.5 km. The sludge contaminated the Prosesspruit River and several other streams. Many other irrigation and drinking water systems were also polluted.

As investigators commence the collection of post-failure evidence for analysing the breach, initial observations point to overtopping as the underlying cause of the failure.[7] Besides destroying over 50 homes, disrupting water supplies and crops, and displacing nearly 400 people, the disaster resulted in the tragic loss of one human life and approximately 900 farm animals owned by at least 25 farmers.[8]

Establishing and maintaining effective monitoring systems for tailings dams is a complex challenge influenced by technical intricacies, site-specific issues, cost constraints, human errors and data management difficulties.[5] Despite



Sources: (A,B): Minerals Council of South Africa[5]; (C) Daily Maverick[6]; (D) Dennis Kelly[6]; reproduced with permission.

Figure 1: (A) An aerial view of the Merriespruit tailings dam after it collapsed; (B) images of the town covered by the sludge; (C) the extent of pollution of the Black Umfolozi River; (D) toxic coal mining waste flowing through Hluhluwe-iMfolozi Game Reserve.

the increasing awareness of potential risks linked to tailings dams, there are significant areas for improvement in detecting and monitoring failures. Monitoring systems often need to be more adequately designed and maintained, and the absence of standardised monitoring guidelines results in inadequate safety measures. Moreover, mining companies rarely share information about their monitoring systems, hindering their effectiveness of evaluation.[9] Communities downstream of these dams are often unaware of the associated risks due to a lack of transparency regarding tailings dam safety.

To address these gaps, increasing investments in monitoring and establishing a transparent reporting system is imperative for accurately assessing and managing these risks. The tailings dam failures observed globally have raised serious questions concerning how government and mining companies deal with the long-term repercussions of 150 years of industrial mining. This study illustrates the potential of using passive seismic interferometry as a tool to image dam walls.

Passive seismic interferometry is a geophysical technique used to estimate properties of the subsurface through the analysis of ambient seismic noise cross-correlation. Alternative approaches to seismic interferometry, such as cross coherence and deconvolution interferometry, are also utilised to derive valuable information about the subsurface environment.[10] This technique has proven to be useful in various applications, including seismic imaging[11], tracking and monitoring groundwater levels[12,13] and exploration of hydrocarbon reservoirs[14]. This technique aims to retrieve the empirical Green's function of the subsurface by using naturally occurring seismic waves generated by various factors, including wind, ocean waves and human activities.

The empirical Green's function serves as an estimate of the Green's function, revealing how a medium, like the Earth's subsurface, responds to a localised energy source. In the field of seismology, the empirical Green's function proves to be a valuable tool, providing essential insights into the acoustic and elastic characteristics of the subsurface.[15,16]

Several studies have successfully demonstrated the possibility of recovering surface waves using seismic interferometry.[15-17] Examining the Earth's subsurface structure encompasses the analysis of surface waves across diverse time intervals. Typically, research in this field has primarily utilised waves with frequencies significantly below 1 Hz to investigate the properties of the crust and upper mantle. However, recent progress in seismic interferometry has broadened the possibilities by extending exploration to frequencies beyond 1 Hz. This advancement has introduced novel opportunities for investigating the subsurface, especially in the range spanning tens to hundreds of meters. Consequently, there is a noticeable trend towards focusing on shorter periods, reflecting an increased interest in studying and understanding the Earth's structure at relatively shallow depths in scientific investigations.[11,18]

Calculating dispersion curves is important in the fields of seismology and geophysics, as they provide crucial insights into the relationship between the seismic wave velocity and frequency. These curves illustrate the variations in the velocity of seismic waves across different frequencies, thereby providing valuable information about the structure of the Earth's subsurface. This approach plays a pivotal role in enhancing our capability to interpret the intricate and dynamic characteristics of the Earth's subsurface. It is also crucial for assessing dam wall stability, especially in locating the phreatic surface and determining soil saturation.[17] Rayleigh waves, generated by P and SV waves interacting with a free surface, have opposite responses to saturation, affecting velocity measurements. Love waves, which are generated when SH waves interact with free surfaces, demonstrate a distinctive decrease in velocity with saturation. This characteristic is the reason why Love waves were chosen for analysis in this study.

Data acquisition

We deployed a total of 20 three-component 14 Hz seismic sensors, each equipped with a sealed gel battery and connected to a three-channel seismic data cube. The synchronisation of time between stations was accomplished using internal GPS modules, and the signal was recorded at a rate of 400 samples per second.

These sensors were positioned along a 95-m survey line, maintaining an interstation spacing of 5 m. To safeguard against adverse environmental conditions, potential animal interference and the risk of contact corrosion from water and soil chemicals, both the data cube and the battery were placed in a sealed plastic box and buried. The sensors were installed to align with north, with their horizontal components oriented east-west in relation to north depicted on the map. The sensors were specifically buried on the tailings dam wall, approximately 1.5 m from the foot of the dam wall. The portion of the dam where the sensors were placed represents only a minor segment of the entire tailings dam at a single elevation. Figure 2A provides an aerial view of the tailings dam where we positioned the stations, while Figure 2B illustrates the distribution of the seismic stations. We planted these stations between cone penetration tests conducted on the tailings dam and above a seepage area at the dam's base.

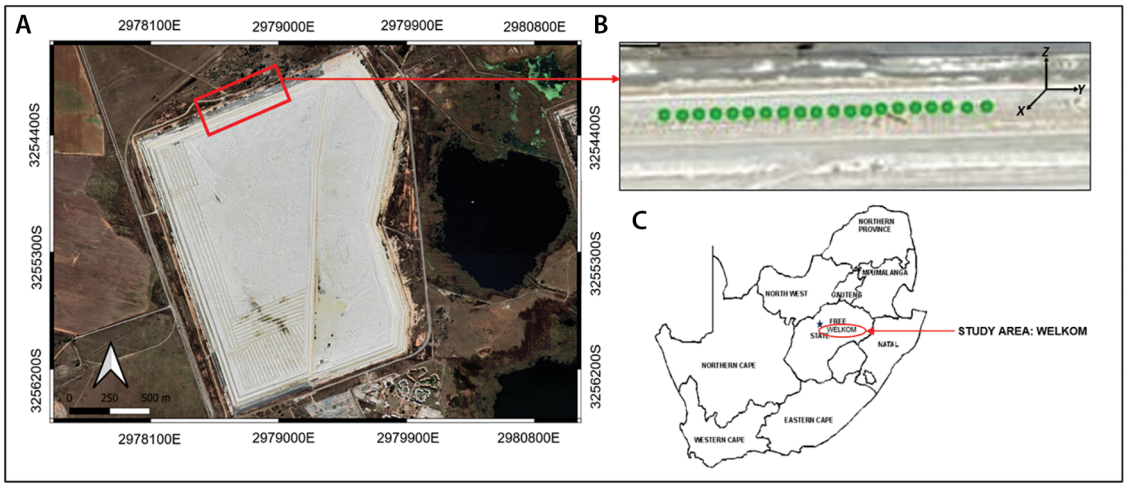
In this project, we follow the processing scheme outlined by Bensen et al.[18] for data processing of a single station. The procedure consists of three main phases: (1) preparation of data from a single station, (2) cross-correlation calculation and temporal stacking and (3) generation of dispersion curves and inversion of ground profiles. For this project, we developed a MATLAB code for the first two phases of the data processing procedure.

Data processing

The ambient seismic noise recordings or data acquired from the stations require pre-processing before immediate use, as the coherent broadband ambient noise needs to be accentuated prior to cross-correlation.

Ambient seismic noise, known as microseisms, emerges in the 2- to 20-second period range due to standing waves in the ocean, serving as a valuable source of information about oceanic processes. Microseisms generated in lakes, observed within the 0.5- to 2-second period range, share potential generation mechanisms with single- and double-frequency microseisms. This information aids in expanding our understanding of seismic phenomena in lacustrine environments.[19] Additionally, microtremors, characterised by shorter periods, typically less than 1 second, are linked to human activities. The frequencies observed in both microseisms and microtremors encompass the frequency band of body and surface waves, providing a holistic view of natural and human-induced seismic influences.[20,21]

Spectrograms function as visual representations that depict the distribution of frequencies throughout the entirety of a recording. Essentially, they offer a comprehensive overview of the frequency composition present in the recorded signal. Particularly noteworthy in this context is the recognition of distinct and prominent peaks surpassing the 50 Hz threshold, as depicted in Figure 3. These high peaks are



Source: Google Earth (created using the free and open source QGIS).

Figure 2: (A) Aerial view of the tailings dam where the seismic stations were planted; (B) a zoomed in picture of Figure 2A showing the distribution of the seismic stations on the tailings dam; (C) outline map of South Africa, the red area indicates the Free State province where Welkom is located.

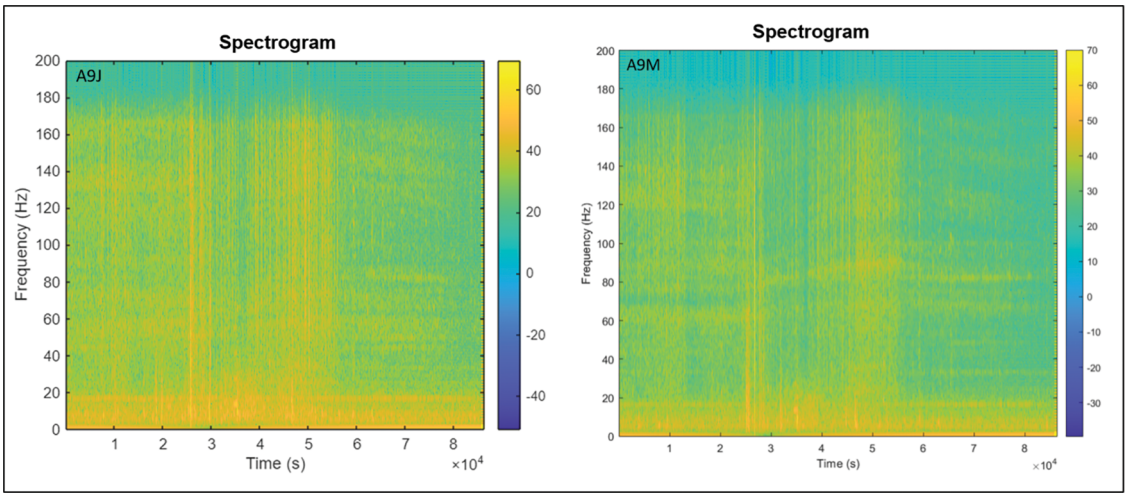


Figure 3: Spectrogram of stations A9J and A9M.

specifically linked to anthropogenic effects, signifying the occurrence of human-generated activities or disturbances within the recorded data.

The spectrogram illustrates temporal variations over time. High frequencies are predominantly observed between 3 and 6 hours, aligning with the commencement of mining operations at the processing plant of the dam being investigated. The heightened frequency during this period is likely associated with human activities. After calculating the spectrograms, the data underwent band-pass filtering to remove high frequencies above 50 Hz before normalisation was applied.

We segmented the data into 10-minute intervals to provide a more detailed view of the signal (Figure 4). This step is crucial for achieving a precise analysis of events and spikes that might not be readily apparent in the original signal.

Temporal normalisation in the time domain assumes significant importance during the preparation of a single station. This reduces the influence of correlated earthquake signals, instrument irregularities and non-stationary noise sources close to the stations.

In this study, we implemented the water-level normalisation method, as explained by Bensen et al.[18] Water-level normalisation is a method commonly used in ambient noise tomography to account for variations in noise levels across different stations or time periods. This normalisation technique involves dividing the seismic data by an estimate of the noise level to achieve a more consistent representation of the ambient noise field. This technique enhances the reliability of seismic data by reducing the influence of non-seismic factors, such as environmental conditions or instrument characteristics, thereby improving the quality of seismic analyses and interpretations.

We applied spectral whitening on the data in the final step of this data processing phase. Ambient noise spectra typically show non-uniformity in the frequency domain, requiring the application of whitening to flatten the spectrum. This step aims to increase the bandwidth of ambient noise while reducing the impact of persistent, spatially isolated noise sources on the signals.

The spectral whitening process involves creating a MATLAB function called 'Whitening'. This function generates a flat Fourier spectrum for a given signal initially exhibiting non-uniformity across a range of frequencies within a defined frequency band. This process includes applying a Hann window to the signal, followed by Fourier transformation, magnitude normalisation and inverse Fourier transformation. Figure 5 illustrates the amplitude spectra of the waveform before whitening, displaying non-uniformity with multiple peaks of varying amplitudes.

Phase 2 of the data processing begins with computation of cross-correlation between all station pairs to obtain empirical Green's functions. The correlation is computed across all components, including north-north, northeast-east, east-east and east-north. Cross-correlation was performed on 10-minute segments between all pairs of sensors. Following computation, the cross-correlations were stacked by summation of the sensor pairs. In this stack, each trace represents the cumulative sum of correlation pairs derived from various sensor pairs.

This summative approach enhances the signal-to-noise ratio and highlights coherent features present in the ambient noise field. The correlation functions pertaining to the vertical component depict estimations of Raleigh waves traversing between sensors, while those of the horizontal transverse signify Love waves. As previously mentioned, the investigation of Love waves is deemed more suitable in this setting. Figure 6 illustrates each trace represents the sum of all the correlation pairs plotted as a function of interstations distance. The cross-correlations are two-sided time functions, encompassing positive and negative time coordinates representing waveforms travelling in opposite directions between stations.

We computed and stored cross-correlation functions for lag times spanning from −2 to 2 seconds. This calculation was determined considering the maximum interstation distance of 95 m (distinct from 85 m) and the slowest anticipated surface wave velocity, and the determined time-series length for the cross-correlation functions is 1.52 seconds. Despite the deployment of 20 stations in the field, it is noted that data from two stations were deemed inadequate for analysis, potentially due to cessation of recording during field operations. Consequently, the data were omitted from the processing and analysis conducted in this study.

The correlation functions of the Love waves show arrival fronts for both positive and negative lag times, related to the velocity of surface wave propagation, and this corresponds to the propagation velocity of the surface waves.

The cross-correlation functions show significant asymmetry in amplitude and spectral content. The observed asymmetry indicates variations between the source and the distance from the source in the radial directions extending outward from the stations.

After completing the computation of daily cross-correlations and the stacking process, we measured phase speeds as functions of the period using the resulting waveform data. The primary objective of this step is to analyse how the velocity of the Love waves changes as a function of frequency.

To calculate the dispersion curves, we employed a phase delay and stack method. In order to create a spatial map of shear wave velocity, consecutive sensors are grouped, and the average dispersion curve for the group is calculated by phase delaying and stacking the correlation functions for each trial velocity at a given frequency. Then, we filtered each stacked correlation function within frequency bands ranging from

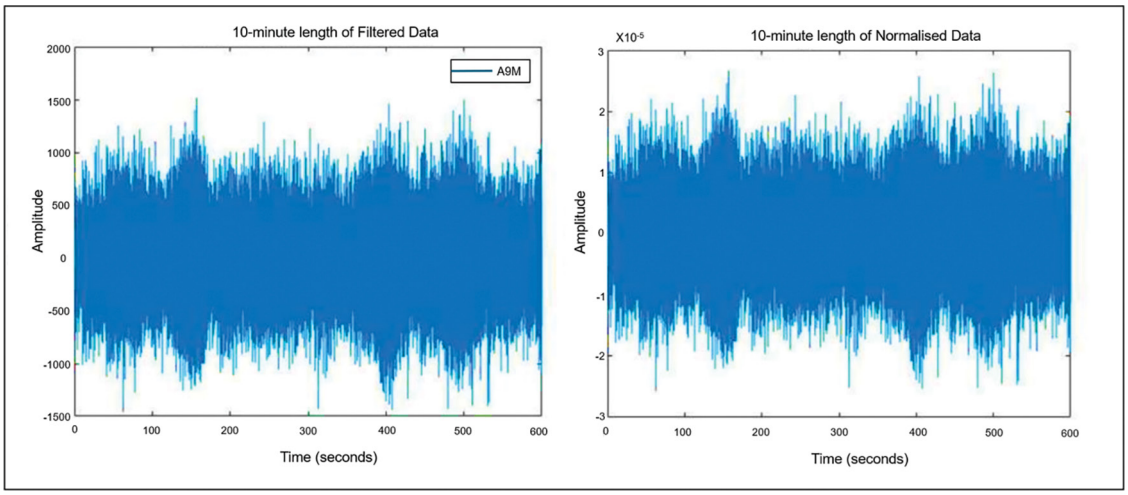


Figure 4: (left) Data that were filtered before being normalised for time using the running absolute mean technique; (right) filtered data after temporal normalisation.

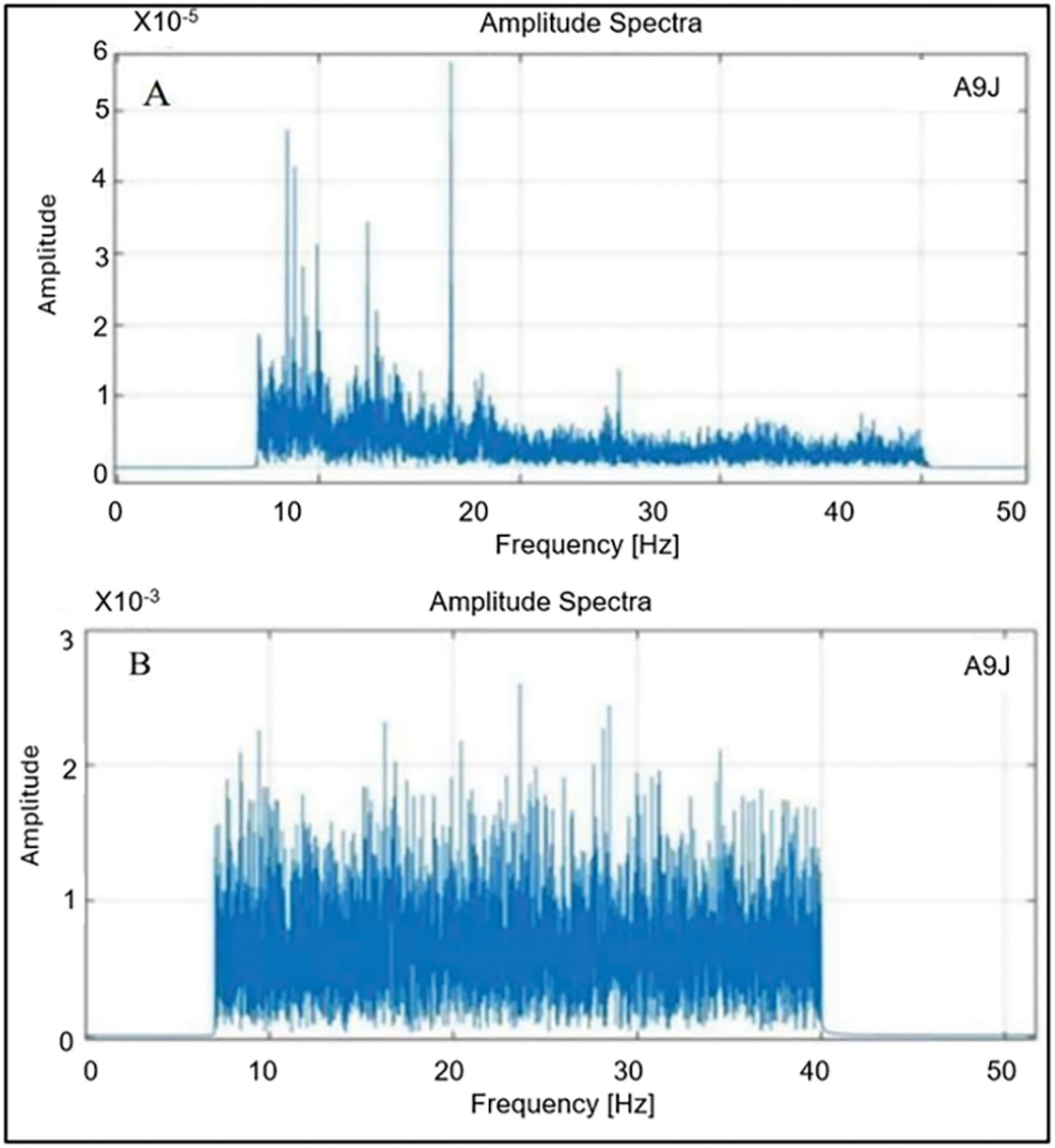


Figure 5: (A) An example of the amplitude spectrum of station A9J before spectral whitening and (B) amplitude spectrum after spectral whitening is applied.

7 Hz to 14 Hz and identified the strongest arrival time for each frequency. Since we knew the distances between the seismic stations, we derived the phase velocity by dividing the distance between the sensors by the selected arrival time. We stacked the curves on each other for multiple pairs of sensors to enhance the signal-to-noise ratio. Figure 7 illustrates the resulting dispersion curves for these sensor pairs.

We calculated average dispersion curves for various velocities at specific frequencies by grouping sensor pairs. The resulting curves consistently demonstrated a common trend, with lower frequencies exhibiting a high velocity range of 138–160 m/s. Some curves displayed relatively high velocities between 10 Hz and 15 Hz, but at frequencies exceeding 20 Hz, certain curves showed a decrease in relative velocity. Anomalous peaks observed in dispersion curves with velocities exceeding 160 m/s may be attributed to the non-uniformity of the ambient noise sources. Variations in the spatial distribution or characteristics of these noise sources within the study area can give rise to localised peaks in dispersion curves.

According to Ivanov et al.[22], surface waves are responsible for approximately 70% of total seismic energy. They are characterised by dispersion, meaning that the different wavelengths have different penetration depths and travel at different velocities. The density and elasticity of the medium influences the speed at which these waves propagate. Longer-period waves are more sensitive to greater depths and penetrate deeper into the Earth's surface and also travel faster generally, as noted by Lin et al.[23]

Surface waves propagate in multiple modes, each with a dispersion curve representing a different velocity range. Higher modes emerge with velocity ranges greater than those of the fundamental mode, with the lowest velocity range. When surface waves propagate through layered horizontal media, they do so multimodally. Different modes of vibration appear at frequencies higher than a specific cut-off frequency. Each mode has its unique propagation velocity, consistently increasing from the fundamental to the higher modes.

In heterogeneous media, higher modes of surface waves emerge due to the constructive interference waves undergoing multiple reflections at layer interfaces.[24] However, detecting these modes is challenging because the energy carried by different overtones is not uniform. The energy distribution depends on frequency, causing a mode to dominate in one frequency band while negligible in another.[24]

We used a wave field transformation procedure explained by Luo et al.[25] to image the various modes and transform the multi-channel record, thereby identifying the dispersion curves based on their altered energy distribution.

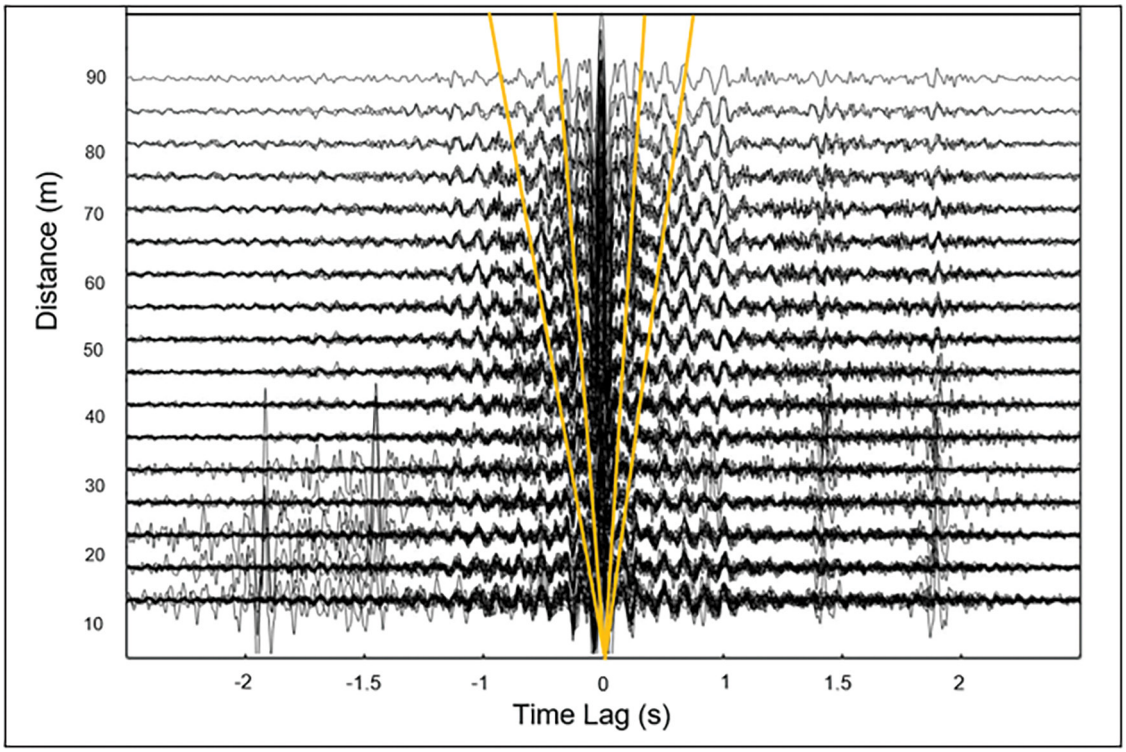


Figure 6: The filtered and normalised data were cross-correlated. The cross-correlation functions are shown against the interstation distance. Both positive and negative lag periods show distinct arrival fronts.

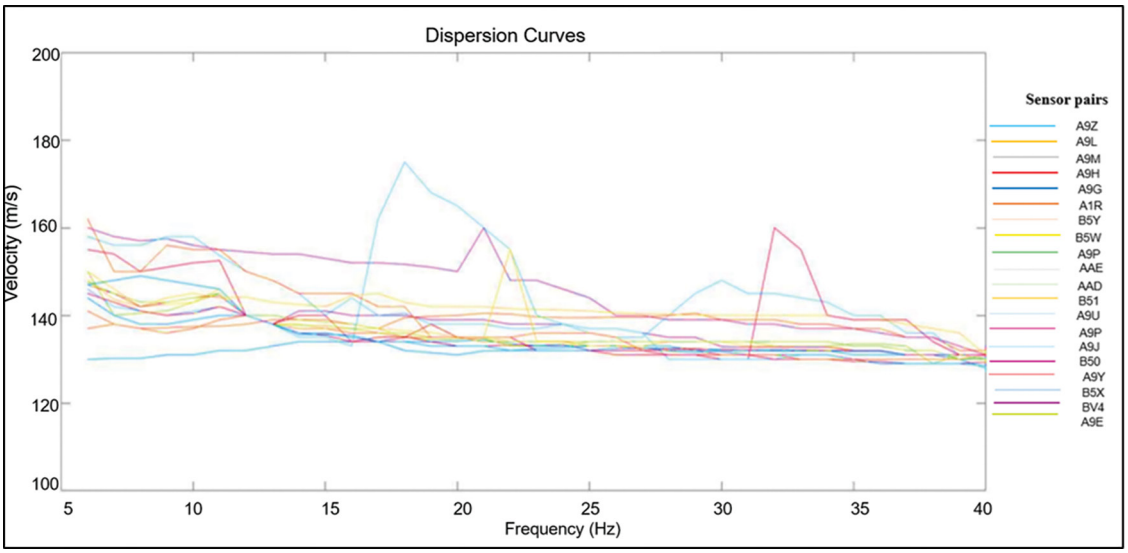


Figure 7: Diagram of the computed dispersion curves (phase velocity plotted against frequency). The plotted lines represent the average dispersion curve for the different sensor pairs.

We employed a Fast Fourier transform to shift the data into the frequency domain, which included phase-shifting the cross-correlation functions. We determined the relative energy for a particular frequency to obtain the slant-stacked amplitude for each set of virtual pairs. The technique included conducting a slant stack of filtered noise cross-correlation traces to create an image in the period-velocity domain. We filtered the traces and stacked them across various velocities at a particular frequency. The stack's amplitude enhances the correct velocity while suppressing others that do not align coherently. We then plotted the summed amplitude in the frequency domain to generate the dispersion curves.[25]

To determine the shear-wave velocity profile with depth, an improved neighbourhood algorithm described by Wathelet[26] was employed to invert the dispersion curves. The velocity changes with depth in the following manner: from the surface down to a depth of 4 m, the velocity spans a range of 240–164 m/s; between 4 m and 14 m, the velocity maintains a constant value of 164 m/s and then decreases to 158 m/s at a depth of 15 m (Figure 8). This decrease in velocity could result from increased saturation in this part of the dam wall.

The inversion analyses were conducted using the Dinver module within the open-source software Geopsy explained by Wathelet[26]. Shear wave velocities generally exhibit low values, with none surpassing 250 m/s at shallow depths between 1 m and 3 m. The average shear wave velocity at a depth of 20 m is 120 m/s. The presence of high velocity anomalies in the theoretical dispersion curve may stem from pre-existing anomalies in the data, which were incorporated into the inversion for analysis.

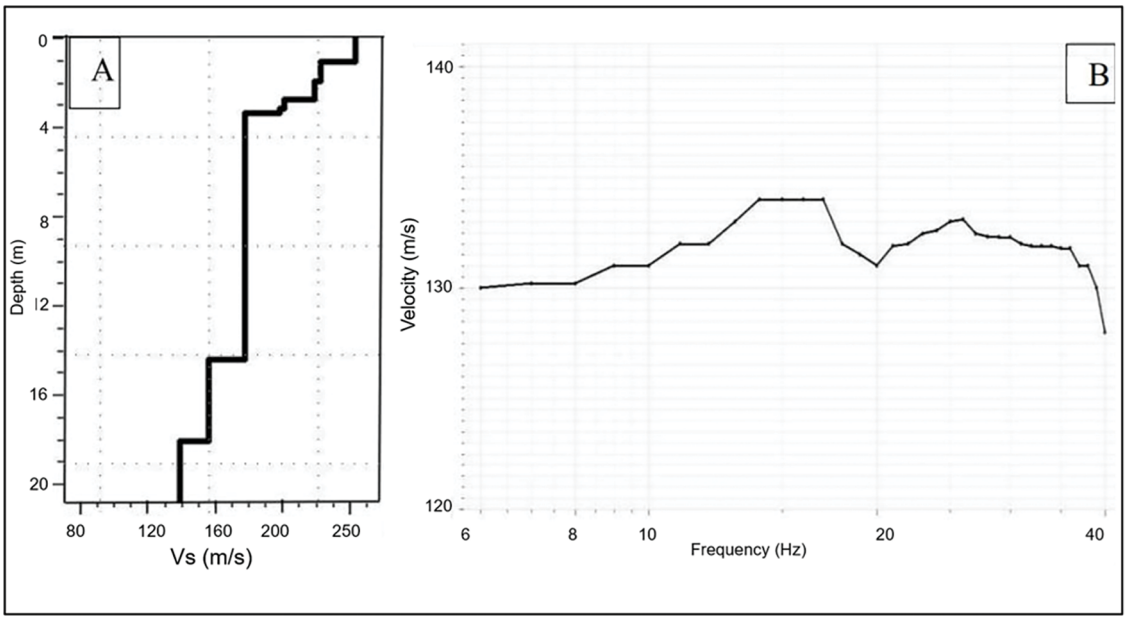


Figure 8: (A) 1D inverted profile and (B) theoretical dispersion curve that corresponds to the velocity model.

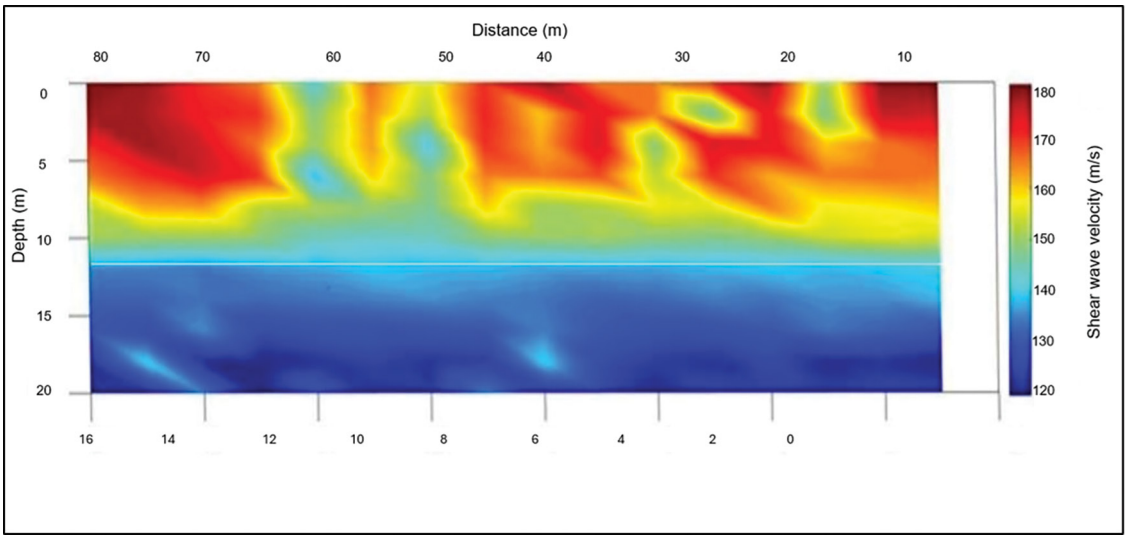


Figure 9: Cross-section of the shear wave velocity profile over 85 m of the section. The phreatic surface is represented by the low-velocity zone underneath the white line. The numbers at the bottom of the figure indicate geophone positions. The profile shows increasing distance from the right to the left and was plotted in this way to be consistent with the order in which the stations were planted. The stations were planted from the left to the right on the tailing dam wall.

The S-wave ground profile indicates higher velocities in the shallow portion (0–9 m). Following the inversion of all 30 dispersion curves, we assigned the resulting S-wave velocity profiles to the centroid of the sensor group used for constructing the shear wave velocity cross-section.

The final phase of the data processing approach involves creating a psuedo 2D shear wave velocity profile to generate a cross-section of surface and depth shear-wave velocities (Vs) (Figure 9). This is achieved by inverting all dispersion curves and implementing a bilinear interpolation method to create a 2D velocity grid data set. Subsequently, we utilised this data set to create a cross-section where different colour codes represent varying velocities.

Bilinear interpolation, a 2D interpolation technique, interpolates values along one axis and then along the perpendicular axis. It extends linear interpolation and is applied to interpolate functions with two variables on a rectangular 2D grid.

The horizontal resolution of the map is influenced by two field variables: the length of the receiver spread and the acquisition period. The virtual source spread length establishes the theoretical lower limit for spatial resolution, and any lateral dimension of a Vs structure smaller than this will not receive accurate resolution in the resulting Vs map. Spatial resolution can also be affected by processing accuracy, as each 1D Vs profile may contain errors arising from inaccuracies in the dispersion-inversion procedure's analysis.

Since the accuracy of stacking velocity significantly impacts the quality of seismic images, we employed interpolation to provide a more precise estimation of the stacking velocity. The initial step in determining stacking velocity involves interpolating time between picks at velocity analysis points. The subsequent step entails spatial interpolation to estimate stacking velocities at positions between the known velocity analysis points.

The upper layers at depths of 0 to 5 m display lateral variations in S-wave velocity, likely attributable to differences in soil compaction. Higher-velocity values may be associated with dense sand, while lower-velocity values may result from loose sand or soft clay.

A low-velocity zone that is close to the surface is detectable at 30 m, 50 m and 60 m along the profile, suggesting the possible presence of very loose sand or very soft clay in this region. Additionally, the S-wave profile shows a distinct low-velocity zone situated approximately 10 m below the surface. Two plumes with lower velocities ranging from 133 m/s to 140 m/s are observed between 50 m and 65 m near the array's core. This observation may indicate that the phreatic surface lies closer to the ground's surface.

Between distances of 70 m and 80 m, a high-velocity zone is discernible on the profile, suggesting the possible existence of dense sand. It is worth noting that the survey line closely corresponds to the area where the photograph providing evidence of seepage was taken (Figure 10).

Conclusion

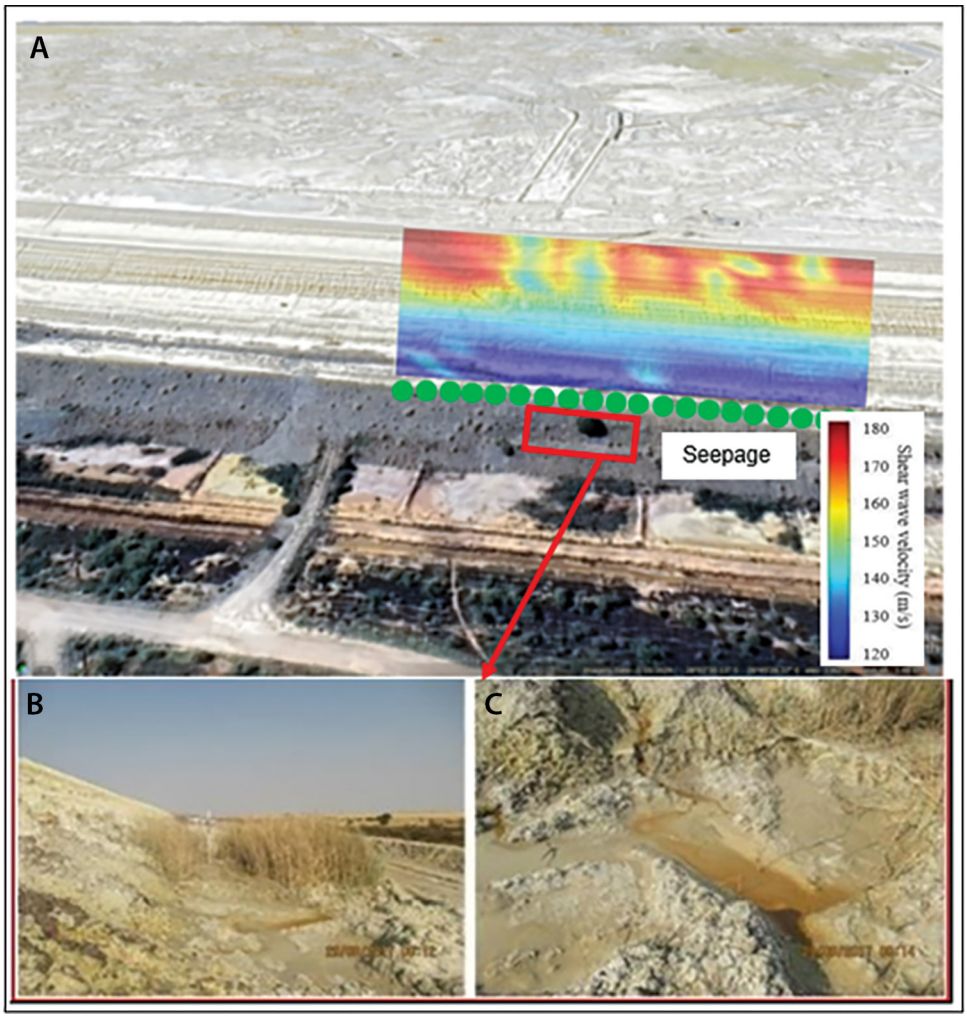
Tailings dams play a critical role in storing mining waste, but they come with the inherent risk of failure, which can lead to severe environmental and human consequences. Traditional methods for assessing dam stability are resource-intensive and expensive, often relying on borehole drilling and physical measurements. Surprisingly, seismic methods, which offer continuous imaging of dam interiors, are underutilised in this context.

There is an urgent need for research and the development of innovative monitoring methods to track the evolving behaviour of these structures over time. Recent advancements in seismic interferometry have provided a way to use ambient noise to create virtual seismic sources at sensor locations, offering a cost-effective solution for monitoring subtle subsurface changes within tailings dams using standard passive micro-seismic monitoring equipment.

Our study aimed to employ ambient seismic noise to image the internal structure of a section of a tailings dam experiencing increased seepage. To accomplish this, we performed cross-correlation on seismic noise recorded from horizontal transverse geophone components, allowing us to estimate Love wave virtual source signals across the sensor array. We generated dispersion curves between different sensor pairs using these virtual source signals. To obtain 1D shear wave velocity profiles, we applied an improved neighbourhood algorithm to invert the dispersion curves.

Furthermore, we utilised virtual source signals to create dispersion curves between sensor pairs and employed the improved neighbourhood algorithm to invert these curves, resulting in 1D shear wave velocity profiles. We interpolated the profiles to construct a 2D shear wave cross-section along the survey line.

The resulting cross-section revealed a low-velocity zone approximately 12 m below the surface. Intriguingly, it also detected similar low-velocity zones at depths of 15, 50 and 60 m along the array, all positioned between 2 m and 10 m below the surface. These findings strongly suggest that the phreatic surface is notably closer to the surface in this specific region.



Source: (A) Google Earth.

Figure 10: The cross-section of the shear wave velocity profile overlain on the aerial image of the tailings dam to correlate with where the stations were deployed. (B) and (C) Pictures that were taken at the foot of the tailings dam. The images show indications of visible surface water and increased plant growth at the dam's base.

In summary, our study showcases the considerable potential of ambient noise tomography as a powerful tool for monitoring the structural integrity of tailings dams. By providing detailed subsurface images, this approach has the potential to identify potential risks and ultimately enhance the safety of these critical structures.

Funding

AEON provided financial support through its NRF-funded Iphakade programme.

Data availability

The data supporting the results of this study are available upon request to the corresponding author.

Acknowledgements

We thank Dr Lucian Bezuidenhout for his invaluable help with the seismic deployment and Dr Gerrit Olivier for assisting with a guide for the processing workflow.

Declarations

We have no competing interests to declare. We have no AI or LLM use to declare.

Authors' contributions

T.K.: Methodology, data collection, data analysis, data curation, writing – the initial draft. M.D.: Conceptualisation, writing – revisions, student supervision, project leadership, funding acquisition. Both authors read and approved the final manuscript.

References

1. Wang X, Zhan H, Wang J, Li P. The stability of tailings dams under dry-wet cycles: A case study in Luonan, China. Water. 2018;10(8):1–11. https://doi.org/10.3390/w10081048
2. Davies MP, Martin PE. Upstream constructed tailings dams – A review of the basics. In: Proceedings of the Tailings and Mine Waste. Boca Raton, FL: CRC Press; 2000. p. 3–15. https://doi.org/10.1201/9781003078579-2
3. Karombo T. South Africa has the world's highest number of environmentally dangerous tailing dams [webpage on the Internet]. c2020 [cited 2023 Mar 20]. Available from: https://qz.com/africa/1786297/south-africa-has-most-environmentally-dangerous-tailingdams
4. Stark TD, Moya L, Lin J. Rates and causes of tailings dam failures. Advances in Civil Eng. 2022;2022:1–21. https://doi.org/10.1155/2022/7895880
5. Minerals Council of South Africa. Merriespruit, 22 February 1994 [webpage on the Internet]. c2023 [cited 2024 Jul 02]. Available from: https://www.mineralscouncil.org.za/industry-news/we-care-we-remember/324-we-remember-merriespruit
6. Carnie T. River turns black after coal mine dam collapse next to rural communities and Hluhluwe-iMfolozi game reserve. Daily Maverick. 2022 January 11. Available from: https://www.dailymaverick.co.za/article/2022-01-11-river-turns-black-after-coal-mine-dam-collapse-next-to-rural-communities-and-hluhluwe-imfolozi-game-reserve/
7. Hansen K. South African town covered in mining waste after dam collapse. SciTechDaily. 2022 October 23. Available from: https://scitechdaily.com/south-african-town-covered-in-mining-waste-after-dam-collapse/
8. Njilo N. Free State mine dam burst floods kill three; four people critically injured. Daily Maverick. 2022 September 12. Available from: https://www.dailymaverick.co.za/article/2022-09-12-free-state-mine-dam-burst-floods-kill-three-four-people-critically-injured/
9. Clarkson L, Williams D, Seppälä J. Real-time monitoring of tailings dams. Georisk: Assess Manag Risk Eng Syst Geohazards. 2021;15(2):113–127. https://doi.org/10.1080/17499518.2020.1740280
10. Cheng F, Xia J, Xu Y, Xu Z, Pan Y. A new passive seismic method based on seismic interferometry and multichannel analysis of surface waves. J Appl Geophys. 2015;117:126–135. https://doi.org/10.1016/j.jappgeo.2015.04.005
11. Yang Y, Ritzwoller M. Characteristics of ambient seismic noise as a source for surface wave tomography. Geochem Geophys Geosyst. 2008;9(2):1–18. https://doi.org/10.1029/2007GC001814
12. Mao S, Lecointre A, van der Hilst RD, Campillo M. Space-time monitoring of groundwater fluctuations with passive seismic interferometry. Nat Commun. 2022;13:1–9. https://doi.org/10.1038/s41467-022-32194-3
13. Clements T, Denolle M. Tracking groundwater levels using the ambient seismic field. Geophys Res Lett. 2018;45(13):1–11. https://doi.org/10.1029/2018GL077706
14. Lehujeur M, Vergne J, Schmittbuhl J, Zigone D, Le Chenadec A. Reservoir imaging using ambient noise correlation from a dense seismic network. J Geophys Res Solid Earth. 2018;123(8):6671–6686. https://doi.org/10.1029/2018JB015440
15. Picozzi M, Parolai S, Bindi D, Strollo A. Characterization of shallow geology by high frequency seismic noise tomography. Geophys J Int. 2009;176(1):164–174. https://doi.org/10.1111/j.1365-246X.2008.03966.x
16. Draganov D, Wapenaar K, Thorbecke J. Seismic interferometry: Reconstructing the earth's reflection response. Geophysics. 2006;71:SI61–SI70. https://doi.org/10.1190/1.2209947
17. Carcione JM, Hans HB. Rock physics of geopressure and prediction of abnormal pore fluid pressure using seismic data. CSEG Recorder. 2002;27:9–32.
18. Bensen GD, Ritzwoller MH, Barmin MP, Levshin AL, Lin F, Moschetti MP, et al. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. Geophys J Int. 2007;169(3):1239–1260. https://doi.org/10.1111/j.1365-246X.2007.03374.x
19. Xu Y, Koper KD, Burlacu R. Lakes as a source of short-period (0.5-2 s) microseisms: Microseisms from lakes. J Geophys Res Solid Earth. 2017;122:1–16. https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017JB014808
20. McNamara D, Boaz R. Visualization of the seismic ambient noise spectrum. In: Nakata N, Gualtieri L, Fichtner A, editors. Seismic ambient noise. Cambridge: Cambridge University Press; 2019. p. 1–18. https://doi.org/10.1017/9781108264808.003
21. Li X, Xu Y, Xie C, Sun S. Global characteristics of ambient seismic noise. J Seismol. 2022;26:343–358. https://doi.org/10.1007/s10950-021-10071-8
22. Ivanov J, Park CB, Miller RD, Xia J. Modal separation before dispersion curve extraction by MASW method. In: Proceedings of the Symposium on the Application of Geophysics to Engineering and Environmental Problems; 2001 March 04–07; Denver, CO, USA. Utrecht: European Association of Geoscientists & Engineers; 2001. p. 1–11. https://doi.org/10.3997/2214-4609-pdb.192.SSM_3
23. Lin FC, Ritzwoller MH, Townend J, Bannister S, Savage MK. Ambient noise Rayleigh wave tomography of New Zealand. Geophys J Int. 2007;170(2), Art. #649666. https://doi.org/10.1111/j.1365-246X.2007.03414.x
24. Foti S, Hollender F, Garofalo F, Albarello D, Asten M, Bard PY, et al. Guidelines for the good practice of surface wave analysis: A product of the InterPACIFIC project. Bull Earthquake Eng. 2018;16:2367–2420. https://doi.org/10.1007/s10518-017-0206-7
25. Luo Y, Xia J, Miller RD, Xu Y, Liu J, Liu Q. Rayleigh-wave mode separation by high-resolution linear Radon transform. Geophys J Int. 2009;179(1):254–264. https://doi.org/10.1111/j.1365-246X.2009.04277.x
26. Wathelet M. An improved neighborhood algorithm: Parameter conditions and dynamic scaling. Geophys Res Lett. 2008;35(9), Art. #L09301. https://doi.org/10.1029/2008GL033256

AUTHORS:
Simphiwe Nundze[1]
Adeniyi Ogunlaja[2]
Kirstie Eastwood[3]
Melusi Thwala[1,4]
Paula Melariri[1]

AFFILIATIONS:
[1]Environmental Health Department, Nelson Mandela University, Gqeberha, South Africa
[2]Chemistry Department, Nelson Mandela University, Gqeberha, South Africa
[3]Stastical Consultation Unit, Nelson Mandela University, Gqeberha, South Africa
[4]Science Advisory and Strategic Partnerships, Academy of Science of South Africa, Pretoria, South Africa

CORRESPONDENCE TO:
Paula Melariri

EMAIL:
paulaezinne.melariri@mandela.ac.za

DATES:
Received: 29 Apr. 2023
Revised: 02 July 2024
Accepted: 12 Sep. 2024
Published: 04 Dec. 2024

HOW TO CITE:
Nundze S, Ogunlaja A, Eastwood K, Thwala M, Melariri P. Investigation of nanomaterial and hazardous emissions at coal-fired power stations in Mpumalanga, South Africa. S Afr J Sci. 2024;120(11/12), Art. #16062. https://doi.org/10.17159/sajs.2024/16062

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITORS:
Michael Inggs
Thywill Dzogbewu

KEYWORDS:
coal-fired power stations, soil, emissions, nanoparticle

FUNDING:
Health and Welfare Sector Education and Training Authority

Investigation of nanomaterial and hazardous emissions at coal-fired power stations in Mpumalanga, South Africa

Coal-fired power stations remain the main source of electricity generation in South Africa. The combustion of coal creates fly ash and slag, and increases emissions of particulate matter, which is composed of nano-sized materials. In this study, we investigated nanoparticle emissions from coal-fired power stations. Soil samples were collected at 500 m and 1 km radii from Matla and Kriel power stations. The soil samples were examined using scanning electron microscopy (SEM), transmission electron microscopy (TEM) and X-ray fluorescence (XRF) spectrometry. SEM images confirmed polydisperse particles in the form of semi-spherical, semi-oval, irregular-shaped and amorphous particles in dust and soil samples. The particle size range was 4–150 nm. Carbon sheet–metal oxide composites of As, Zn, Cu, Cr, Ni and V were observed. We found that coal-fired power stations are a potential source of nano-pollution, pointing to elevated human and environmental exposure around such sites. Currently, there are no environmental limits for nanomaterials due to the lack of robust risk assessment; however, this study suggests that coal-fired power stations may be hotspots that could be used as priority cases to examine the environmental implications of nano-pollution.

Significance:

Coal-fired power stations are a potential source of nano-pollution, pointing to elevated human and environmental exposure around such sites. Currently, there are no environmental limits for nanomaterials due to the lack of robust risk assessment; however, our findings suggest that coal-fired power stations may be hotspots that could be used as priority cases to examine the environmental implications of nano-pollution.

Introduction

Approximately 30% of the global energy need is met through coal, with the remainder from oil, natural gas, nuclear energy, hydroelectricity, and renewable sources.[1] South Africa is the fifth largest producer and sixth largest consumer of coal globally[2]; about 90% of its electricity is from coal-fired power stations supplied by the Electricity Supply Commission (Eskom)[3]. Eskom operates 13 coal-fired power stations with an estimated 39 342 Megawatt (MW), excluding Kusile and Medupi power stations that are yet to operate at full capacity. An additional 46 540 MW capacity comes from nuclear (1940 MW), hydropower (2732 MW), open cycle gas turbine (2426 MW) and wind farming (100 MW).[4]

Coal combustion and processing lead to the formation of particulate matter (PM) and ultrafine particles (UFPs) that are released into the environment due to vaporisation of inorganic substances in coal.[5] These vaporised particles form a variety of nanoparticles (NPs) as by-products through nucleation, which then coagulate and aggregate when they condense to form accumulated mode aerosols.[6] Generally, nanomaterials (NMs) have at least one size dimension within the 1–100 nm range, and they can be in any phase of matter. NMs can be natural (e.g. ocean sprays, fine sand, dust, volcanoes, biological matter) or anthropogenic (incidental and engineered). Incidental NMs are by-products of human activities, for instance, motor vehicle emissions, mining, coal-based power stations and fires.[7,8] Engineered nanomaterials (ENMs) are specifically designed and synthesised for applications; examples include quantum dots, carbon nanotubes, gold NPs and fullerenes, among others.[9]

Burning of coal emits various airborne pollutants, for instance, trace metals, mercury (Hg), sulfur dioxide (SO_2), sulfur trioxide (SO_3), nitrogen oxides (NO_x), condensable PM, and radioactive nucleoids, which can be persistent in the environment.[10] However, the emission properties depend on the type and quality of coal.[11] The elemental composition of NPs emitted from coal combustion includes proportions of aluminium (Al), carbon (C), calcium (Ca), iron (Fe), magnesium (Mg), sodium (Na), sulfur (S), silica (Si), arsenic (As), chromium (Cr), lead (Pb), vanadium (V) and zinc (Zn) among others.[6]

Approximately 25% of coal combustion emissions are associated with trace metals[12], and relatively more volatile trace elements, such as Hg and Se, are released in abundance as compared to As.[13] Concentrations of Ni, Zn, Cr, Cd and Pb notably above background natural soil levels have been recorded globally in coal-fired power station surroundings, including in South Africa.[14] Coal burning emits substantial amounts of $PM_{2.5}$ µm or UFPs through the process of mineral transformation at high temperatures, causing negative environmental implications, such as the release of fly ash particulates, acid rain, and emission of greenhouse gases (GHG) like carbon dioxide (CO_2).[15,16] A study conducted at the Coal Terminal in Richards Bay (South Africa) found that coal dust harms the local aquatic (mainly photosynthesis inhibition in mangroves) and terrestrial ecosystems.[17]

Attention has been given to PM_{10} and smaller particulates because of their relatively large surface area, which is associated with enhanced potential to induce human health effects such as lung cancer, heart diseases and asthma.[18] When considering the health risks associated with NPs and UFPs, the more hazardous forms tend to be the smaller sized, which have higher adsorptive and absorptive potential and can, with relative ease, reach organs such as the kidneys and brain.[19]

The aim of the current study was to investigate NP emissions in coal-fired power stations and further examine compliance witha the South African soil quality regulations for trace metals. The study objectives were to:

1. Determine levels and physicochemical characteristics of NPs in soil samples collected at various points at the Matla and Kriel power stations.
2. Assess the measured values of arsenic, lead, chromium, copper, zinc, vanadium and nickel against the acceptable levels as per the National Norms and Standards for the Remediation of Contaminated Land and Soil Quality in the Republic of South Africa.

Materials and methods

Description of the study area

The study was conducted at Eskom's Matla and Kriel coal-fired power stations (hereafter Matla and Kriel) (Figure 1). The proximity of the two power stations was ideal for practical purposes as they are both located within the Emalahleni Local Municipality in Mpumalanga province. The two coal-fired power stations are located in the Highveld Priority Area (HPA), which is associated with poor air quality due to intense mining, the concentration of coal-fired power stations, and industrial and agricultural activity. The HPA covers 31 106 km^2 that includes parts of Gauteng and Mpumalanga provinces.[20]

Matla is approximately 10 km outside the town of Kriel, and when it was completed in 1979, it was the largest coal-fired power station in the southern hemisphere, being among the first to be supplied with coal from a fully mechanised coal mine. Matla generates 3000 MW and has approximately 694 employees.[21]

Kriel has a generation capacity of 3600 MW; it was the first of the giant coal-fired power stations that were commissioned in the 1980s. Kriel was designed for an operating lifespan of 30 years, but it has since been extended to 50–60 years. A total of 3800 tons of coal per hour can be transported by conveyor from a nearby colliery to the power station, resulting in the consumption of approximately 1 150 000 tons of coal per month. There are approximately 700 employees (excluding students and contractors) at this station.[22] Geographic information system (GIS) mapping was done using QGIS software[23] to show NP distribution levels across Matla and Kriel power stations.

Sample collection

Soil and dust samples were collected at Matla and Kriel to determine the physicochemical properties and levels of NPs and heavy metal concentrations following procedures described by the US Environmental Protection Agency Laboratory Services & Applied Science Division.[25] At both power stations, 30 samples were collected: 10 soil samples at a 1 km radius, 10 soil samples at a 500 m radius and 10 dust samples from windows, desks and filing cabinets of offices in the power stations (Figures 2 and 3).

Soil samples

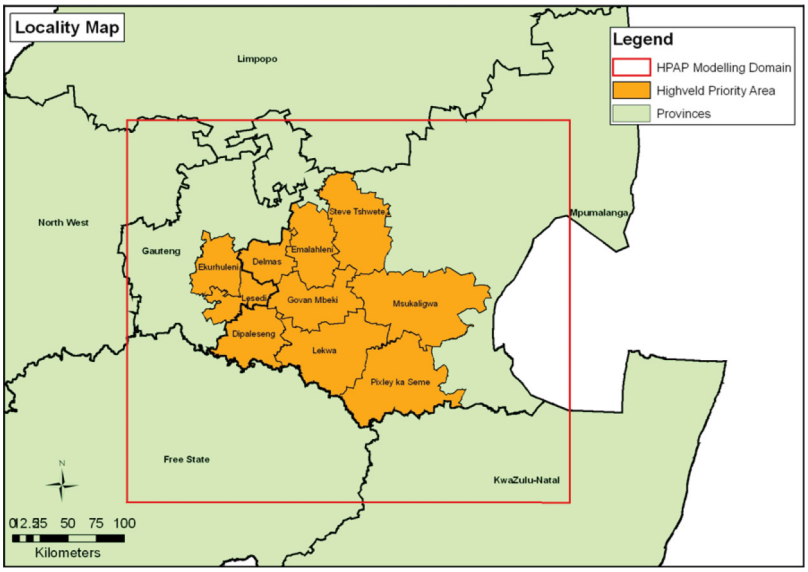
Soil samples at each power station were collected using a metal scoop at a depth of 1.2 cm. The obtained samples were placed in paper trays and thoroughly mixed, and foreign materials such as stones, roots, gravels and pebbles were removed. Each soil sample was placed in a ziplock plastic container and labelled according to the sampling point, the name of the power station and radius. To prevent the cross-contamination of samples, wipes and distilled water were used to disinfect the metal scoop until the soil was completely removed before taking another sample.

Dust sample collection

Dust samples were collected from windows, desks and filing cabinets of offices at the power stations using the ASTM E 1728-03 procedure, a standard practice for the collection of settled dust samples using wipe sampling methods for subsequent hazardous material determination. The dust samples were collected from surfaces of 30 X 30 cm to standardise the size of the collection surface areas. Ten dust samples were collected from each power station. The sampling area was wiped from the left to the right at either corner to the furthest. Wipes were folded to keep the sampling side in and to prevent loss of dust samples collected. A second wiping was from the top to the bottom repeating the same procedure, and then the wipes were folded in half again, with the sample side inside. Lastly, wiping was around the perimeter of the sampling area, and the wipes were folded as previously described. For each sample, wipes were stored in a sterile plastic container, sealed with a lid and labelled.

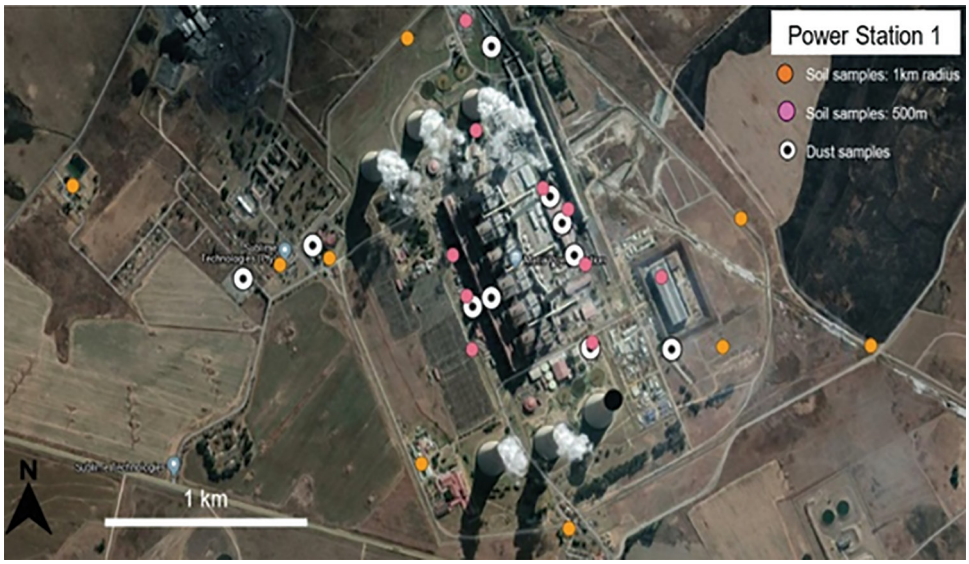
Heavy metal analytes

Seven heavy metals, namely As, Cr, Cu, V, Ni, Pb and Zn, were prioritised for analysis due to their notable adverse environmental health effects.[26] Their concentrations were assessed against the norms and standards for the remediation of contaminated land and soil quality of the *National Environmental Management: Waste Act, 2008 (Act No. 59 of 2008)*.[27] Based on the known land usage (i.e. in the HPA), the application of commercial/industrial soil screening values (SSVs) in the norms and standards for contaminated land and soils was justified, as the study examined human exposure through direct and indirect routes.



Source: South African Department of Environmental Affairs[24] (non-commercial re-use permitted)

Figure 1: An illustrative map of the Highveld Priority Area in South Africa.



Source: Map created using open-source QGIS

Figure 2: Soil and dust sample collection points at Matla Power Station.

Source: Map created using open-source QGIS

Figure 3: Soil and dust sample collection points at Kriel Power Station.

Analytical instruments

Transmission electron microscopy

The transmission electron microscope (TEM; Carl Zeiss Libra 120 equipped with an energy-dispersive X-ray analysis (EDX) detector and Gatan Crystorage, Germany) was used to characterise the particle size and morphology of the soil samples. All soil samples collected were placed in methanol. The solution was then spread out using a plastic Pasteur pipette onto a 3.05 mm carbon-coated copper grid and left to dry at room temperature. The samples were mounted on a TEM carbon-coated specimen holder using a fine-point tweezer. The images were captured using the embedded self-imaging system with a MegaView III digital camera.

Scanning electron microscopy

The scanning electron microscopy (SEM) analysis of soil samples was undertaken using the TESCAN Vega TS 5136LM (TESCAN, Czech Republic) operating at 20 kV at a working distance of 20 mm. The SEM was coupled with EDX for elemental analysis. A double-sided carbon adhesive tape was used for sample attachment on sample stubs. The samples were sprinkled evenly but lightly on an SEM sample stub with double-sided sticky tape and a hand blower used to blow away the loose particles. They were mounted on the SEM stubs using Storkbill forceps to avoid unintentional damage to the sample.

Data analysis

Data analysis was performed using the IBM SPSS Statistics version 27. Descriptive and inferential statistics were used to analyse and describe the data and included the Student's *t*-test and chi-square (χ^2) tests (Table 1). Hypothesis testing was conducted with a significance level of $\alpha = 0.05$; significant associations were further interrogated using Cohen's d for tests involving sample means, and Cramer's V for tests involving frequencies to determine practical significance. Cohen's values larger than 0.20 and Cramer's V larger than 0.10 indicate practically significant results.

Table 1: Independent samples test

	Levene's test for equality of variances		*t*-test for equality of means		
	F	**Significance**	***t***	**d.f.**	***p*-value**
Arsenic	3.593	0.074	1.779	18	0.092
Total chromium	1.783	0.198	0.304	18	0.765
Copper	0.802	0.383	0.224	17	0.826
Lead	3.813	0.067	0.621	18	0.542
Nickel	0.651	0.430	-1.082	18	0.294
Vanadium	0.239	0.631	-3.522	18	0.002
Zinc	0.000	0.986	-0.803	18	0.432

Results and discussion

Determination of NPs

SEM Analysis

The SEM images of soil samples collected at the 500 m and 1 km radius are provided in Figure 4. In the dust samples from the power station offices (Figure 4B and C) and soil samples (Figure 4A, D, E and F), individual NPs within agglomerates could be identified and some were embedded within the organic matrix, indicating that NPs are hardly found as individual particles but rather as agglomerates. The particles were polydisperse in terms of size and morphology. There was a mixture of semi-spherical, semi-oval and irregular-shaped particles in the soil samples. Due to agglomeration influence, there was size polydispersity. The interaction occurring between particles may be influenced by environmental conditions and physicochemical characteristics of the particles, which determine the size dynamics of NPs.[28] Generally, the soil samples were predominantly aggregated, possibly indicative of swelling due to organic combustion as observed by bubbles produced around the sampled coal-based materials.[29]

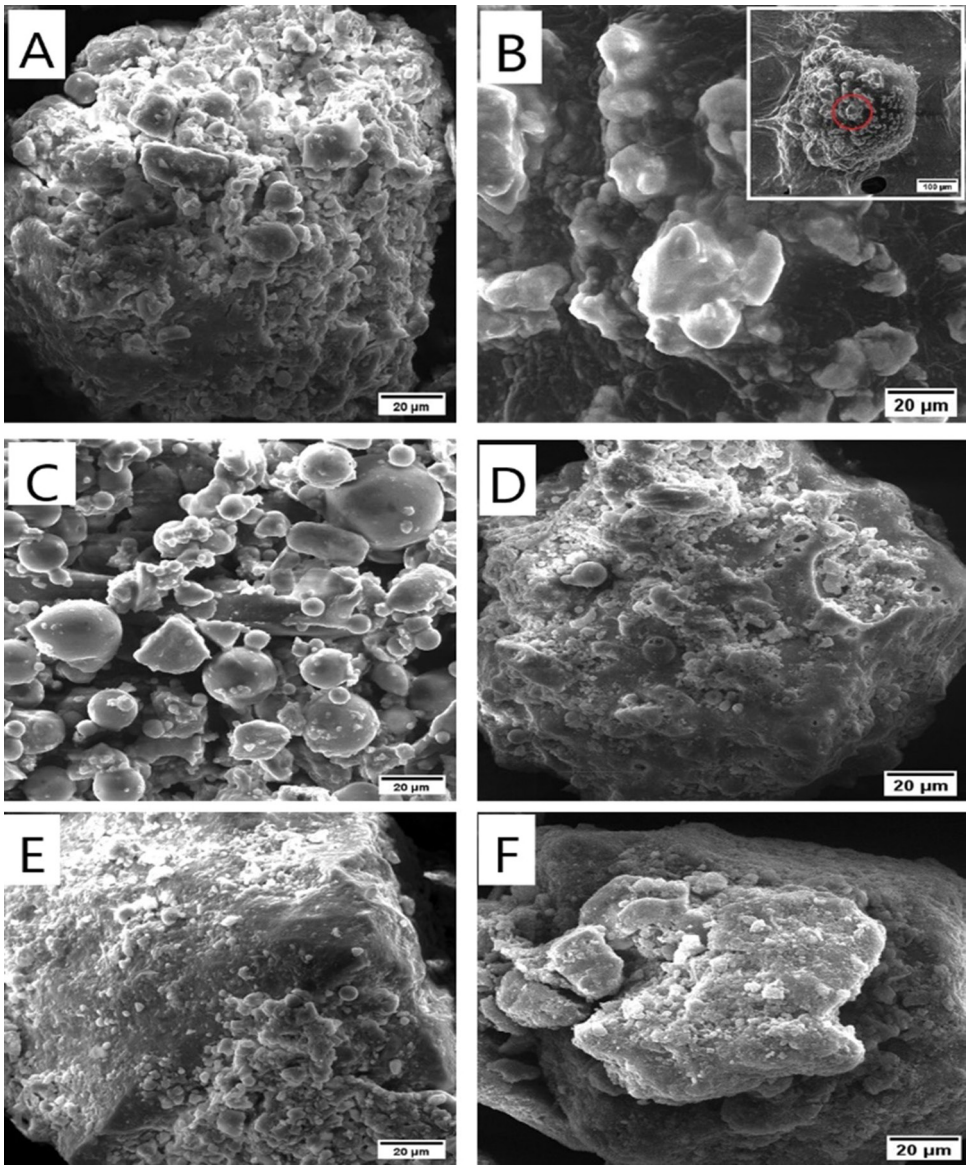


Figure 4: SEM images of dust and soil samples collected at Matla and Kriel: (A) soil samples collected outside the Mechanical Maintenance Department, 500 m, (B) dust sample collected at environmental offices, (C) dust sample collected at slurry offices, (D) soil samples collected outside the outage offices, 500 m, (E) soil samples collected at contract area west, 1 km and (F) soil sample collected at the coal stock yard, 1 km.

Furthermore, the semi-spherical morphology is distinct for coal combusted at high temperature through the process of decomposition, nucleation, coagulation and condensation of vaporised material.[29] The degree of solid organic matter was more prominent on samples collected within the 500 m radius for both Matla and Kriel, compared to samples from the 1 km radius. Such findings indicate carbonic signatures are stronger close to the point of emission, that is, the power station.

TEM Analysis

The TEM analysis enabled the identification of individual primary particles within agglomerates in all samples (Figure 5A–G). Layers of carbon sheets and graphitic carbon-based NMs were observed on the images with no dispersed particles within the images, meaning that the nanocomposites were mainly non-metal-based. Metal-based NPs embedded within carbon-based sheets were observed and were sized approximately 6–9 nm (Figure 5E–G). The average NP size was 4 nm and 1 nm in Matla and Kriel dust samples, respectively. For soil samples, the average NP size was 15 nm at Matla and 1 nm at Kriel. The average length of NPs in dust was 14 nm and 2 nm in Matla and Kriel, respectively. In soils, the average length of NPs was 113 nm at Matla and 3 nm at Kriel.

Size is a major factor determining the reactivity of NPs, including toxicity potential. The smaller the size of the particle, the greater the surface-area-to-volume ratio, and thus the higher the reactivity and toxicity potential. For instance, UFPs with sizes ranging from 12 nm to 20 nm (Figure 5H) have the ability to penetrate the alveolar lining and enter the lungs at a rapid rate compared to larger counterparts.[30,31]

Assessment of samples

Dust samples

The results of heavy metal concentrations (mg/kg) in dust samples from Matla and Kriel are presented in Table 2. There was generally a concentration variation within each power station, and there was no uniform trend when comparing concentrations between the power stations. As, Cr, Cu, V, Ni, Pb and Zn were found in all dust samples collected at both Matla and Kriel. The concentration of As at Matla (30 404 mg/kg) and Kriel (13 551 mg/kg) was higher than those of the other heavy metals, whereas Ni at both sites was the least abundant. A study conducted at Baotou city, China, showed that human activities have a significant impact on emissions of heavy metals. High concentration values for As, Co and C were observed in areas where steel smelting and thermal power stations are located.[32]

The South African SSV of 150 mg/kg for As in industrial/commercial areas was exceeded in dust samples from both power stations. The considerably high extent of the As SSV exceedance levels suggests the need for precautionary measures (e.g. frequent cleaning of surfaces and utilisation of necessary protective personal equipment), and establishment of monitoring to establish trends in order to safeguard human health. It is noteworthy that As was only detected in dust samples and not in soil samples for both power stations, suggesting that emissions were probably indoors, although such a source(s) was not investigated further as this extended beyond the original scope of the study. As is widespread but rarely found naturally and average concentration levels could be 2 mg/kg in the natural environment[33], hence the recorded values considerably above natural background levels strengthen credence for an anthropogenic source. Pesticide application, waste incineration and smelting are among the known As sources in the environment.[34]

The study focused on total Cr using X-ray fluorescence (XRF) to determine the elemental composition of soil. Cr average concentrations were 751 and 599 mg/kg, respectively, for Matla and Kriel dust samples; these were above the limit for South African industrial/commercial areas of 40 mg/kg. This may be due to the large amount of Cr compounds (e.g. $FeCr_2O_4$) present close to coal fire plants.[35] All other heavy metals (Cu, Pb, Ni, V, Zn) examined in both power stations fell below the respective SSVs as outlined in South African norms and standards. Average As, Cr, Cu and Pb in dust samples were relatively higher at Matla than at Kriel; however, Ni, V and Zn were relatively higher at Kriel.

Soil samples

Soil samples from 500 m radius

Cr, Cu, V, Ni, Pb and Zn were detected in soil samples from both Matla and Kriel; however, As was not detected (Table 3). Cr exists in a variety of oxidation states, the most common of which are Cr (III) and Cr (VI). The

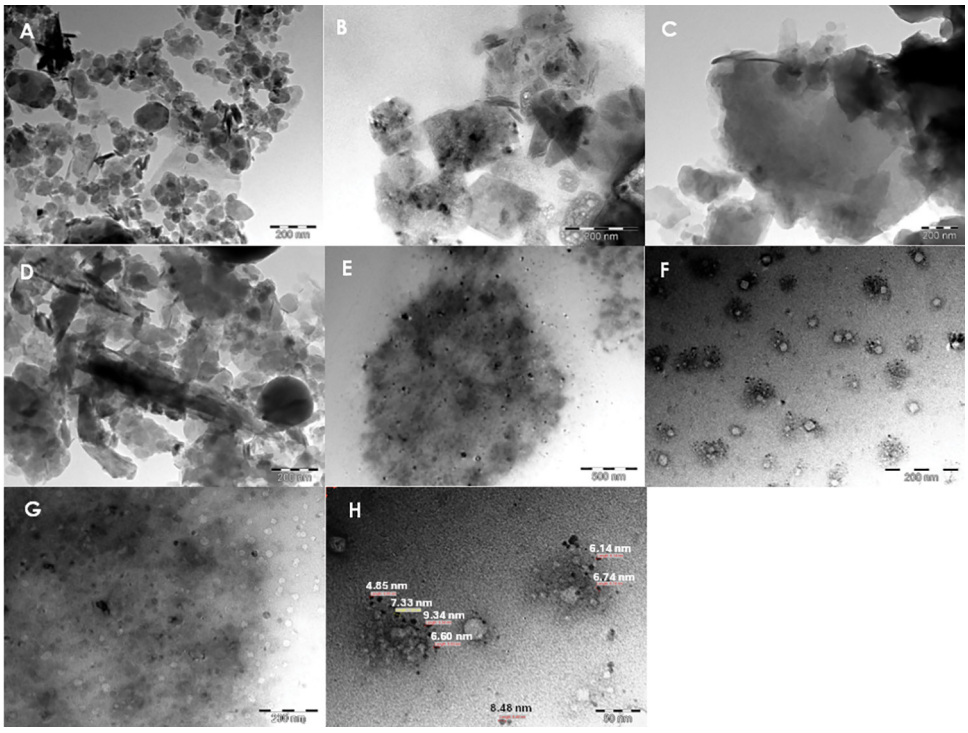


Figure 5: TEM images of (A) soil sample collected outside the outage offices, 500 m, (B) soil sample collected at the coal stock yard, 1 km, (C) soil sample collected outside the Mechanical Maintenance Department, 500 m, (D) TEM image showing spherical oval and rod-like particles, (E) dust sample collected at slurry offices, (F) soil sample collected at contract area west, 1 km, (G) dust sample collected at the environmental offices and (H) particle size distribution for metal-based nanoparticles.

difference between these two species is critical because Cr (III) is often regarded as benign or even needed for some biological activities, but Cr (VI) is extremely toxic and carcinogenic. The average Cr concentrations were 328 mg/kg and 815 mg/kg at Matla and Kriel, respectively, which exceed the SSVs for metals and organics per South African norms and standards (40 mg/kg).

The prevalence of heavy metals in mining operations can be explained by their natural occurrence in the earth's crust, soil, air and water. According to Kalagbor and Kiadum[36], when Ni is present in tiny amounts, several enzymes become active. It plays a role in fat metabolism and serves as a biocatalyst necessary for body pigmentation. Higher doses are thought to be carcinogenic, can irritate the skin, and cause damage to the heart and liver. On the other hand, Cu is a naturally occurring metal that builds up in plants and animals and is a micronutrient that is necessary for overall health. However, excessive concentrations can have negative consequences, such as nausea, diarrhoea, stomach cramps, and irritation of the nose and eyes. Brain tumours and liver cancer may potentially be related to excess Cu.[36] Zn is regarded to be rather non-toxic, particularly when taken orally. Zn toxicosis has been associated with symptoms such as vomiting, bloody urine, liver failure, renal failure and anaemia. It has been known to give similar symptoms to Pb poisoning and is easily misdiagnosed. The average soil concentrations for Cu Pb, Ni, V and Zn at both Matla and Kriel were below the respective SSVs (Table 3 and Figure 6).

Pb has no health benefits of any sort. It is biotoxic and has major consequences, including teratogenicity. Pb poisoning inhibits haemoglobin synthesis and causes kidney, joint, reproductive and cardiovascular problems as well as long-term harm to the central and peripheral nervous systems.[36] Concentrations of Pb at Matla (43 mg/kg) and Kriel (60 mg/kg) were below the South African allowable limit of 19 000 mg/kg for industrial/commercial areas. Natural background Pb concentrations in surface soils can be 3–65.8 mg/kg.[37] Pb was found in 27 of the 30 samples collected at Kriel and 20 of the 30 samples collected at Matla. Agricultural material, metallurgic industries, waste disposal and automotive fuels are some of the known sources of Pb.[37-39] The road network around Matla and Kriel, which commonly carries trucks and employee transport, possibly contributes to the detected Pb concentrations. Although leaded fuels are being phased out in South Africa, there are vehicles that still use leaded petrol or diesel.[40]

Table 2: Descriptive statistics for dust samples collected at Matla and Kriel (mg/kg)

Site	Heavy metals	Mean	Median	Standard deviation	Range	%> South African norms and standards	South African norms and standards
Matla	As	30 404.30	22 018.50	28 026.51	96 546	100%	150
	Total Cr	751.10	266.00	1521.95	5000	80%	40
	Cu	114.00	122.00	90.40	264	80%	19 000
	Pb	158.20	49.50	360.17	1174	60%	1900
	Ni	44.30	31.50	37.87	98	30%	10 000
	V	40.40	0	56.63	153	10%	2600
	Zn	824.30	675.50	656.48	2157	80%	150 000
Kriel	As	13 550.90	9921.50	10 584.737	31 470	100%	150
	Total Cr	598.80	481.00	439.096	1263	90%	40
	Cu	101.40	55.00	145.430	460	60%	19 000
	Pb	87.10	80.50	34.411	130	100%	1900
	Ni	62.70	58.50	38.178	137	10%	10 000
	V	132.10	140.00	59.773	213	40%	2600
	Zn	1080.30	987.00	765.219	2609	100%	150 000

Table 3: Average metal concentrations in soil at Matla and Kriel, with respective limits at 500 m and 1 km radius

Heavy metal	Matla soil average 500 m (mg/kg)	Matla soil average 1 km (mg/kg)	Kriel soil average 500 m (mg/kg)	Kriel soil average 1 km (mg/kg)	Allowable limit of heavy metals in soil for industrial/commercial areas (mg/kg)
					South African national norms and standards
As	–	–	–	–	150
Total Cr	328	558	815	212	40
Cu	132	106	58	62	19 000
Pb	43	28	60	40	1900
Ni	82	100	62	71	10 000
V	148	136	164	222	2600
Zn	192	105	261	152	150 000

Soil samples from 1 km radius

Between the two power stations, Cr, Cu and Ni average concentrations were relatively higher at Matla, whereas concentrations of Pb, V and Zn were higher at Kriel (Table 3); however, all were below the respective SSVs.

Cr was high at both power stations (Matla: 558 mg/kg, Kriel: 212 mg/kg) against the 40 mg/kg limit for South Africa and was detected in 29 of the 30 samples from Kriel and 24 of the 30 samples from Matla. Cr is primarily released from coal combustion and waste slurry.[10] Natural in origin and extensively utilised in industrial operations, it is regarded as one of the most dangerous heavy metals. The exceeded soil contamination limit values for South Africa demonstrated that Cr in coal is closely associated with ash-forming minerals.[35] High Cr in coal ash may also stem from the grinding media or as a result of stainless-steel erosion of power plant installation.[41] Human exposure to Cr can occur through inhalation and accumulate in the lungs (among other parts); nasal ulcers, skin hypersensitivity and chronic oral effects' have been reported in exposed humans.[42,43]

Average Pb concentrations were 28 and 40 mg/kg at Matla and Kriel, respectively; this was below the South African set limit of 1900 mg/kg. The South African limits for Ni and V were not exceeded at Matla and

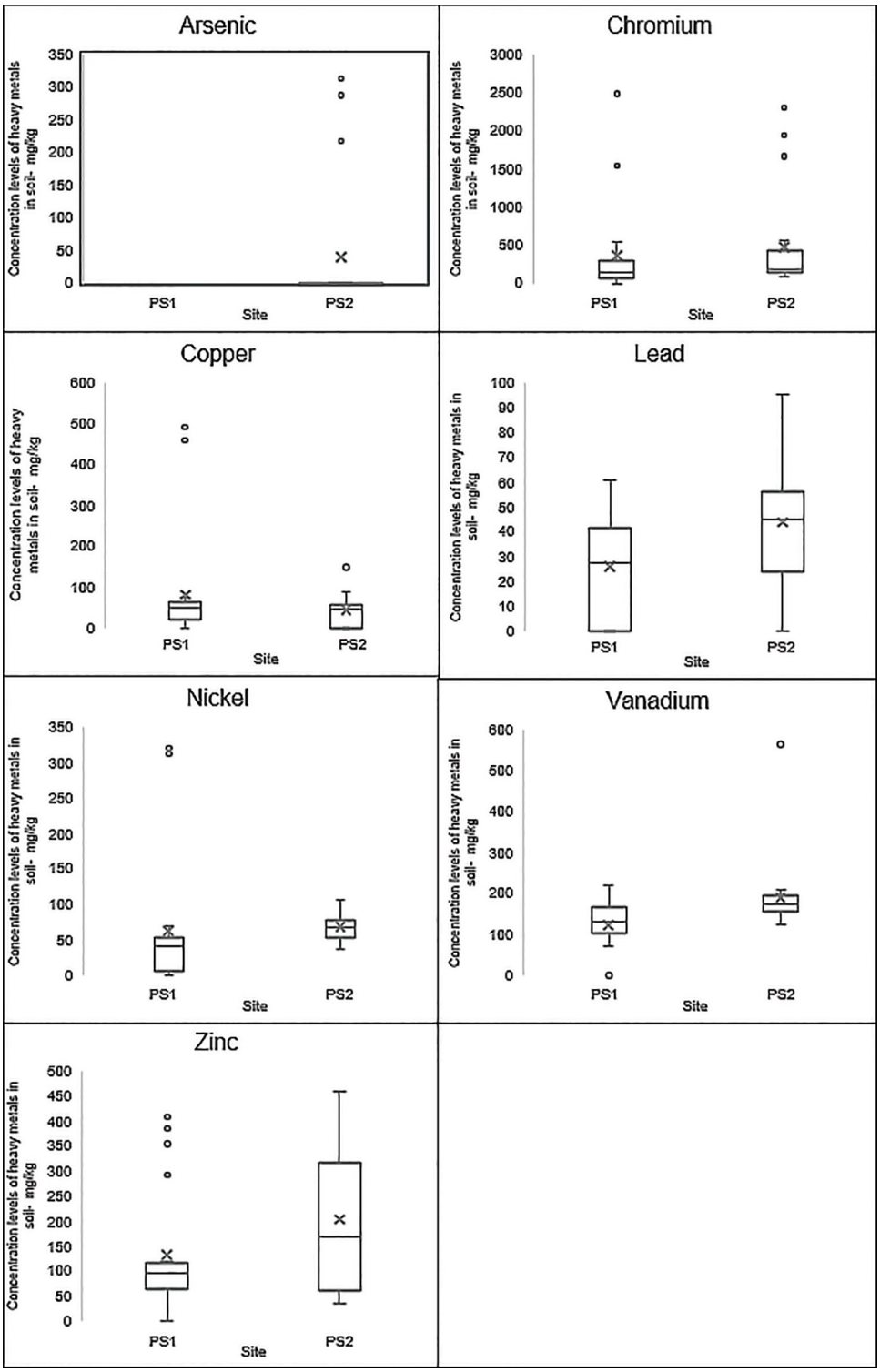


Figure 6: Graphical presentation of descriptive statistical parameters of analysed soil concentrations from Matla and Kriel. (The x symbols represent mean; • depicts outliers; PS1 = Matla; PS2 = Kriel.)

Kriel. The Zn average concentration was 105 mg/kg and 152 mg/kg at Matla and Kriel, respectively. These averages were below the 150 000 mg/kg guideline for South Africa (Table 3).

In soil samples, the average range for Matla was 2419 mg/kg, and for Kriel, it was 32 149 mg/kg. Additionally, the data pointed to Cr, Cu and Ni at a radius of 1 km at Kriel being higher compared to Matla. On the other hand, the concentration levels of Zn, V and Pb at Kriel were higher at a radius of 1 km compared to those at Matla. In 2015, the power utility applied for exemptions for 13 of its existing power stations as contained in GNR 893 as amended by GNR 1207 (31 October 2018), which were granted.[44] The high concentrations of heavy metals in soil could be a result of the non-compliance of both power stations to minimum emission standards (PM_{50} mg/Nm^3).

Comparison of hazardous material concentrations in study sites

On comparing inter-site differences in concentrations of As, Cr, Pb, Ni, V and Zn, no significant difference between the power stations was observed for As ($p = 0.092$), Cr ($p = 0.765$), Cu ($p = 0.826$), Pb ($p = 0.542$), Ni ($p = 0.294$) and Zn ($p = 0.432$) for both 500 m and 1 km radius. However, for V, the average concentrations varied significantly for both power stations between 500 m and 1 km radius ($p = 0.002$), being relatively higher at 500 m .

Natural V soil concentration can reach 100 mg/kg. Combustion of heavy fuels, especially in coal-fired power stations, refineries, industrial boilers and coal mines are major sources of anthropogenic emissions of V.[45] In the case of Kriel, V emissions could be additionally to a coal mine located 5 km from the power station.

Conclusion and recommendations

We have confirmed the presence of anthropogenically derived NMs in the soil samples that were collected at both power stations. The presence of NPs in the soil confirmed environmental and potential occupational exposure. Currently available studies have generally focused on UFPs, NMs and heavy metals in coal, stockpile and air samples[12,46] and not on exposure assessments in soil and dust samples, as in the current study. In South Africa, coal fly ash has been predominantly the medium of interest[47]; as a result, comparative analysis with other local findings is not possible. Furthermore, comparison with values from other countries would not be suitable due to natural geological differences. In this regard, the current study will support comparative assessment with future local studies.

The study confirmed the presence of Cr at levels that exceeded the South African average metal concentrations in soil. The average levels of Cr in the soil samples from both power plants were higher than what the South African norms and standards permitted, associated with the release of Cr in coal and ash-forming minerals. High Cr in coal ash might also stem from the stainless-steel erosion of power plant installations. Average concentrations of Pb at both power stations were below the South African recommended limit. However, Pb can also be derived from other sources, for instance, emissions from vehicles.

Average concentrations for V and of Matla and Kriel soil samples were below the South African maximum allowable limit. Likewise, Zn concentrations at both power stations did not exceed the South African allowable limit. Furthermore, for Zn and Pb, it was observed that concentrations varied between sampling distances.

The common exceedance of heavy metal limits in soil is a concern for human and environmental health. This concern calls for more stringent pollution control measures at the power stations; however, other activities should not be overlooked. Due to the presence of anthropogenically derived NPs in soil at power stations, such sites present likely hotspots for human and environmental exposure to NMs. Such power stations can be prioritised to establish environmental exposure guidelines for NMs. As part of the corporate health, safety, and environment policies, it is recommended that exposure analysis through dust in work settings be incorporated as part of long-term monitoring in order for the establishment of credible exposure and effects data.

Acknowledgements

We thank Eskom for their knowledge sharing and for providing information without which this publication would not have been possible. We also thank the anonymous reviewers for their constructive inputs.

Funding

We thank the Health and Welfare Sector Education and Training Authority for funding this research project.

Data availability

Data are available on request from the corresponding author.

Declarations

We have no competing interests to declare. We declare no use of AI tools.

Authors' contributions

S.N.: Conceptualisation; methodology; data collection; sample analysis; data analysis; validation; writing – the initial draft; writing – revisions; funding acquisition. A.O.: Conceptualisation; methodology; sample analysis; data analysis; validation; writing – revisions; student supervision; project management. K.E.: Data analysis; validation; data curation; writing – revisions. M.T.: Conceptualisation; methodology; data analysis; validation; writing – revisions; student supervision; project management. P.M.: Conceptualisation; methodology; data analysis; validation; writing – revisions; student supervision; project leadership; project management. All authors read and approved the final manuscript.

References

1. International Energy Agency (IEA). Report extract coal [webpage on the Internet]. c2019 [cited 2020 Aug 30]. Available from: https://www.iea.org/reports/world-energy-outlook-2019/coal#abstract
2. World Coal Association. Coal statistics [document on the Internet]. c2020 [cited 2020 Aug 30]. Available from: https://www.worldcoal.org/file_validate.php?file=WCA_South%20Africa_0.pdf
3. Todd I, McCauley D. Assessing policy barriers to the energy transition in South Africa. Energy Policy. 2021;158, Art. #112529. https://doi.org/10.1016/j.enpol.2021.112529
4. Eskom Fact Sheet. Generation plant mix, revision 19 [document on the Internet]. c2017 [cited 2020 Oct 21]. Available from: https://www.eskom.co.za/AboutElectricity/FactsFigures/Documents/GX0001GenPlantMixRev19.pdf
5. Hochella MF Jr, Mogk DW, Ranville J, Allen IC, Luther GW, Marr LC, et al. Natural, incidental, and engineered nanomaterials and their impacts on the Earth system. Science. 2019;363(6434), eaau8299. https://doi.org/10.1126/science.aau8299
6. Saikia BK, Saikia J, Rabha S, Silva LF, Finkelman R. Ambient nanoparticles/nanominerals and hazardous elements from coal combustion activity: Implications on energy challenges and health hazards. Geosci Front. 2017;9(3):863–875. https://doi.org/10.1016/j.gsf.2017.11.013
7. Lohse S. Nanoparticles are all around us [webpage on the Internet]. c2013 [cited 2021 May 28]. Available from: https://sustainable-nano.com/2013/03/25/nanoparticles-are-all-around-us/#:~:text=Naturally%20occurring%20nanoparticles%20can%20be,than%20their%20naturally%20occurring%20counterparts
8. Ray U. What are the different types of nanoparticles? [webpage on the Internet]. c2018 [cited 2021 May 28]. Available from: https://www.azonano.com/article.aspx?ArticleID=4938
9. Occupational Safety and Health Administration (OSHA). OSHA fact sheet, working safely with nanomaterials [webpage on the Internet]. c2020 [cited 2020 Aug 30]. Available from: www.osha.gov/dsg/nanotechnology/nanotechnology.html
10. Savic D, Nisic D, Malic N, Dragosavljevic Z, Medenica D. Research on power plant ash impact on the quality of soil in Kostolac and Gacko coal basins. Minerals. 2018;8(2):54. https://doi.org/10.3390/min8020054
11. Shahzad BK, Yousaf M. Coal-fired power plants: Emission problems and controlling techniques. J Earth Sci Clim Change. 2017;8:404. https://doi.org/10.4172/2157-7617.1000404

12. Okedeyi OO, Dube S, Awofolu OR, Nindi MM. Assessing the enrichment of heavy metals in surface soil and plant (*Digitaria eriantha*) around coal-fired power plants in South Africa. Environ Sci Pollut Res. 2013;21:4686–4696. https://doi.org/10.1007/s11356-013-2432-0

13. Nalbandian H. Trace element emissions from coal. IEA Clean Coal Centre, 601, CCC/203 [document on the Internet]. c2012 [cited 2021 Jun 27]. Available from: https://usea.org/sites/default/files/092012_Trace%20element%20emissions%20from%20coal_ccc203.pdf

14. Dalton A, Feig GT, Barber K. Trace metal enrichment observed in soils around a coal-fired power plant in South Africa. Clean Air J. 2018;28(2), Art. #1. https://doi.org/10.17159/2410-972x/2018/v28n2a1

15. Lelieveld J, Evans JS, Fnails M, Giannadaki D, Pozzer A. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature. 2015;525:367–371. https://doi.org/10.1038/nature15371

16. Finkelman RB, Wolfe A, Hendryx MS. The future environmental and health impacts of coal. Energy Geosci. 2021;2(2):99–112. https://doi.org/10.1016/j.engeos.2020.11.001

17. De Place E. Coal dust threatens cascadia's water and wildlife. Seattle, WA: Sightline Institute; 2016 [cited 2021 Apr 25]. Available from: https://www.sightline.org/2016/05/12/coal-dust-threatens-cascadias-water-and-wildlife/

18. World Health Organization (WHO). Ambient air pollution: A global assessment of exposure and burden of disease. Geneva: WHO; 2016 [cited 2021 Jun 27]. Available from: https://iris.who.int/bitstream/handle/10665/250141/?sequence=1

19. Mahar BA, Ahmedb IAM, Karloukovski V, MacLaren DA, Foulds PG, Allsop D, et al. Magnetite pollution nanoparticles in the human brain. Proc Natl Acad Sci USA. 2016;113(39):10797–10801. https://doi.org/10.1073/pnas.1605941113

20. Pauw C, Garland RM, Oosthuizen R, Albers P, John J, Wright CY. Air quality and human health among a low-income community in the highveld priority area. Clean Air J. 2011;20(1):12–20. https://doi.org/10.17159/caj/2011/20/1.7180

21. Matla Power Station. Generation plant mix, generation division corporate affairs division [webpage on the Internet]. c2020 [cited 2020 Nov 30]. Available from: https://www.eskom.co.za/Whatweredoing/ElectricityGeneration/PowerStations/Coal/Pages/Matla_Power_Station.aspx

22. Kriel Power Station. Generation plant mix, generation division corporate affairs division [webpage on the Internet]. c2020 [cited 2020 Nov 30]. Available from: https://www.eskom.co.za/Whatweredoing/ElectricityGeneration/PowerStations/Coal/Pages/Kriel_Power_Station.aspx

23. GIS Development Team. Spatial without compromise | QGIS web site. 2021 [cited 2021 Jun 27]. Available from: https://qgis.org/

24. South African Department of Environmental Affairs (DEA). Highveld Priority Area air quality manage-ment plan. Pretoria: DEA; 2011 [cited 2021 Jun 27]. Available from: https://www.gov.za/sites/default/files/gcis_document/201409/35072144.pdf

25. US Environmental Protection Agency. Soil sampling operating procedure [document on the Internet]. c2023 [cited 2023 Jul 01]. Available from: https://www.epa.gov/sites/default/files/2015-06/documents/Soil-Sampling.pdf

26. Holland M. Health impacts of coal-fired power stations in South Africa [document on the Internet]. c2017 [cited 2021 Jun 27]. Available from: https://cer.org.za/wp-content/uploads/2017/04/Annexure-Health-impacts-of-coal-fired-generation-in-South-Africa-310317.pdf

27. South Africa. Notice 467 of 2013, National Environmental Management: Waste Act, 2008 (Act No 59 of 2008): National norms and standards for the remediation of contaminated land and soil quality [document on the Internet]. c2008 [cited 27 Jun 2021]. Available from: https://faolex.fao.org/docs/pdf/saf90450.pdf

28. Tourinho PS, Van Gestel CAM, Lofts S, Svendsen C, Soares AMVM, Loureiro S. Metal-based nanoparticles in soil: Fate, behavior, and effects on soil invertebrates. Env Toxicol Chem. 2012;31(8):1679–1692. https://doi.org/10.1002/etc.1880

29. Lighty JS, Veranth JM, Sarofim AF. Combustion aerosols: Factors governing their size and composition and implications to human health. Air Waste Manag Assoc. 2000;50:1565–1618. https://doi.org/10.1080/10473289.2000.10464197

30. Morabet RE. Effects of outdoor air pollution on human health. In: Reference module in earth systems and environmental sciences. Amsterdam: Elsevier; 2018 https://doi.org/10.1016/B978-0-12-409548-9.11012-7

31. Yang Y, Ruan Z, Wang X, Yang Y, Mason TG, Lin H, et al. Short-term and long-term exposures to fine particulate matter constituents and health: A systematic review and meta-analysis. Environ Pollut. 2019;247:874–882. https://doi.org/10.1016/j.envpol.2018.12.060

32. Han X, Lu X, Qinggeletu, Wu Y. Health risks and contamination levels of heavy metals in dusts from parks and squares of an industrial city in semi-arid area of China. Int J Environ Res Public Health. 2017;14, Art. #886.https://doi.org/10.3390/ijerph14080886

33. Tabelin CB, Igarashi T, Villacorte-Tabelin M, Park I, Opiso EM, Ito M, et al. Arsenic, selenium, boron, lead, cadmium, copper, and zinc in naturally contaminated rocks: A review of their sources, modes of enrichment, mechanisms of release, and mitigation strategies. Sci Total Environ. 2018;645:1522–1553. https://doi.org/10.1016/j.scitotenv.2018.07.103

34. Chung JY, Yu SD, Hong YS. Environmental source of arsenic exposure. J Prev Med Public Health. 2014;47:253–257. https://doi.org/10.3961/jpmph.14.036

35. Li Z, Wang Q, Xiao Z, Fan L, Wang D, Li X, et al. Behaviors of chromium in coal-fired power plants and associated atmospheric emissions in Guizhou, Southwest China. Atmosphere. 2020;11(9), Art. #951. https://doi.org/10.3390/atmos11090951

36. Kalagbor I, Kiadum D. Evaluation of the presence and concentrations of nine heavy metals: Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd and Pb in five crops harvested from farms located in Kpean and Zaakpon in Khana Rivers State Nigeria using AAS and XRF. Res Rev J Chem. 2022;11(6):6–12. Available from: https://www.rroij.com/open-access/evaluation-of-the-presence-and-concentrations-of-nine-heavymetals-cr-mn-fe-co-ni-cu-zn-cd-and-pb-in-five-cropsharvested-.pdf

37. Bakshi S, Banik C, He Z. The impact of heavy metal contamination on soil health. In: Reicosky D, editor. Managing soil health for sustainable agriculture. Vol 2: Monitoring and management. Sawston: Burleigh Dodds Science Publishing Ltd; 2018. p. 63–95. https://doi.org/10.19103/AS.2017.0033.20

38. Hassaan MA, El Nemr A, Madkour FF. Environmental assessment of heavy metal pollution and human health risk. Am J Water Sci Eng. 2016;2(3):14–19. Available from: https://sciencepublishinggroup.com/article/10.11648/j.ajwse.20160203.11

39. Mast L. Coal ash contain lead, arsenic and mercury – and it's mostly unregulated. Environmental engineering - Georgia Institute of Technology [webpage on the Internet]. c2019 [cited 2021 Mar 25]. Available from: https://massivesci.com/articles/coal-ash-water-health-hazard/

40. Zhu C, Tian H, Hao J. Global anthropogenic atmospheric emission inventory of twelve typical hazardous trace elements, 1995–2012. Atmos Environ. 2020;220, Art. #117061. https://doi.org/10.1016/j.atmosenv.2019.117061

41. Wilczyńska-Michalik W, Janusz Dańko J, Michalik M. Characteristics of particulate matter emitted from a coal-fired power plant. Pol J Environ Stud. 2020;29:1411–1420. https://doi.org/10.15244/pjoes/106034

42. Holmes AL, Wise SS, Wise JP Sr. Carcinogenicity of hexavalent chromium. Indian J Med Res. 2008;128:353–372. Available from: https://digitalcommons.usm.maine.edu/cgi/viewcontent.cgi?httpsredir=1&article=1000&context=appmedsci

43. Ray S, Ray MK. Bioremediation of heavy metal toxicity-with special reference to chromium. Al Ameen J Med Sci. 2009;2(2):57–63.

44. Naledzi Environmental Consultants. Application for postponement of the Minimum Emission Standards (MES) compliance timeframes for Eskom's coal and liquid fuel-fired power stations in the Mpumalanga Highveld, Vaal Triangle and Cape Region, Revision 2 [document on the Internet]. c2019 [cited 2021 Mar 03]. Available from: https://www.eskom.co.za/MESapplications/Documents/Annexure%204.2%20%20Final%20IRR_Updated%20Feb%202019%20v2.pdf

45. Guagliardi I, Cicchella D, Rosa R, Buttafuoco G, Ricca N. Geochemical sources of vanadium in soils: Evidences in a southern Italy area. J Geochem Explor. 2018;184:358–364. https://doi.org/10.1016/j.gexplo.2016.11.017

46. Duarte AL, DaBoit K, Oliveira MLS, Teixeira EC, Schneider IL, Silva LFO. Hazardous elements and amorphous nanoparticles in historical estuary coal mining area. Geosci Front. 2019;10:927–939. https://doi.org/10.1016/j.gsf.2018.05.005

47. Akinyemi SA, Gitari WM, Thobakgale R, Petrik LF, Nyakuma BB, Hower JC, et al. Geochemical fractionation of hazardous elements in fresh and drilled weathered South African coal fly ashes. Environ Geochem Health. 2020;42:2771–2788. https://doi.org/10.1007/s10653-019-00511-3

AUTHORS:
Sandee Oster[1]
Jerome P. Reynard[1]
Hayley C. Cawthra[2,3]
Irene Esteban[2,4,5]
Justin Pargeter[6,7]
Erich C. Fisher[2,8,9,10]

AFFILIATIONS:
[1]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa
[2]African Centre for Coastal Palaeoscience, Nelson Mandela University, Gqeberha, South Africa
[3]Minerals and Energy Unit, Council for Geoscience, Cape Town, South Africa
[4]Archaeological and Archaeometric Research Unit of the University of Barcelona (ERAAUB), Department of History and Archaeology, University of Barcelona, Barcelona, Spain
[5]Institute of Archaeology, University of Barcelona, Barcelona, Spain
[6]Department of Anthropology, New York University, New York, New York, USA
[7]Rock Art Research Institute, School of Geography, Archaeology, and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa
[8]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa
[9]Interdisciplinary Center for Archaeology and the Evolution of Human Behavior (ICArEHB), University of Algarve, Faro, Portugal
[10]Institute of Human Origins, School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, USA

CORRESPONDENCE TO:
Jerome Reynard

EMAIL:
Jerome.Reynard@wits.ac.za

DATES:
Received: 13 Dec. 2023
Revised: 01 Sep. 2024
Accepted: 03 Sep. 2024
Published: 04 Dec. 2024

HOW TO CITE:
Oster S, Reynard JP, Cawthra HC, Esteban I, Pargeter J, Fisher EC. Late Pleistocene and Holocene fauna from Waterfall Bluff Rock Shelter, Mpondoland, South Africa. S Afr J Sci. 2024;120(11/12), Art. #17449. https://doi.org/10.17159/sajs.2024/17449

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

Late Pleistocene and Holocene fauna from Waterfall Bluff Rock Shelter, Mpondoland, South Africa

Archaeological deposits from Waterfall Bluff Rock Shelter (Eastern Cape) span from Marine Isotope Stage 3 (~39–29 ka) to the mid-Holocene (~8 ka), showing persistent human occupations. The site's consistent proximity to the shoreline and stable coastline over millennia makes it key for exploring human settlement patterns. This study reports on preliminary results of identifiable fauna from Marine Isotope Stage 3 to the Early Holocene layers at Waterfall Bluff. The identified species may suggest a mosaic environment, although caution is warranted given the small sample size. Furthermore, leopard seal remains were recovered in layers dating to the Last Glacial Maximum. This is the first direct evidence of a leopard seal recovered from Pleistocene and Holocene archaeo-faunal assemblages along the South African coast.

Significance:

Last Glacial Maximum (LGM) sites are rare in southern Africa. Waterfall Bluff in the Eastern Cape shows that human occupation persisted there from Marine Isotope Stage 3 to the mid-Holocene. A leopard seal tooth was identified in the LGM layers, making it the first evidence of this species recovered off South Africa's coast.

Introduction

During the Terminal Pleistocene, coastal archaeological records are scarce across southern Africa due to sea-level fluctuations, which results in a bias towards interglacial periods.[1] Excavations at Waterfall Bluff (WB), located in Mpondoland in the Eastern Cape, have documented repetitive occupations of the rock shelter by hunter-gatherers from the Marine Isotope Stage 3 (~39–29 ka) to the middle Holocene (~8–5 ka), including during the Last Glacial Maximum (LGM) and the glacial/interglacial transition.[1,2] The abundant remains of marine shellfish, fish, mammals and charcoal from coastal taxa demonstrate that Late Pleistocene hunter-gatherers routinely collected various resources from coastal zones.[1-3] Here, we describe a preliminary sample of WB fauna from the Marine Isotope Stage 3 to the Early Holocene to better understand the local palaeoenvironmental and palaeoecological conditions and the implications for hunter-gatherer subsistence.

Regional ecology

Eastern Mpondoland is known for its diverse landscapes, including deeply dissected plateaus extending from the Mthatha River mouth to the Umtamvuna River mouth. The area is currently dominated by sourveld grassland, forest vegetation and bushveld, including the Southern Coastal Forest, Southern Mistbelt Forest and Scarp Forest.[1,4,5]

The region is home to various animal species in different environments. Browsers such as bushbuck (*Tragelaphus scriptus*), blue duiker (*Philantomba monticola*) and grey rhebok (*Pelea capreolus*) are primarily found in environments that provide underbrush, including riverine underbrush, woodlands and coastal bush.[6,7] Grazers such as the southern reedbuck (*Redunca arundinum*), oribi (*Ourebia ourebi*) and bontebok (*Damaliscus pygargus*) prefer grassland environments with short grass for grazing and longer grass for cover.[6,7] Mixed feeders like the eland (*Tragelaphus oryx*) are versatile in their environmental needs, occurring in grassland biomes to woodlands.[6,7]

Site background

Waterfall Bluff (31°26'01, 1" S, 29°49'19.2" E) is located ~24 m above the modern coastline next to the Mlambomkulu River waterfall. The site's lithics assemblage is predominantly hornfels.[1] Botanical records show the presence of all major vegetation types found in the region today from the end of the Pleistocene to the Holocene.[2] These findings imply a complex mosaic of environments that would have supported varying plant and animal resources.

All excavated data have been mapped in 3D with total stations to millimetric accuracy using total stations tied into the Universal Transverse Mercator (UTM) grid system. More than 17 000 artefacts, faunal (terrestrial and marine) and plant remains have been plotted to date. Additional specialist samples for micromorphology, optically stimulated luminescence dating, charcoal, etc. have been similarly mapped in 3D.

The site's stratigraphic sequence follows natural layers consolidated into 'Stratigraphic Aggregates' (StratAggs), which are laterally continuous layers of sediments.[8] StratAggs can be subdivided into sub-Stratigraphic Aggregates (SubAggs), representing discrete anthropogenic, biogenic or geologic events. The Light Brown Coarse Sands (LBCS) are the earliest StratAgg with dates ranging from 37.6 ± 4.2 ka to 12.5 ± 1.2 ka dated via single-grain optically stimulated luminescence.[1] The Shell-Rich Clayey Sands (SRCS) StratAgg overlays the LBCS and is dated using Bayesian14C accelerator mass spectrometry model from 11 000 cal yr BP to ca. 10 500 cal yr BP[1] (Figure 1). For more detailed information about the excavation sequence, see Fisher et al.[1], Esteban et al.[2] and Karkanas et al.[8]

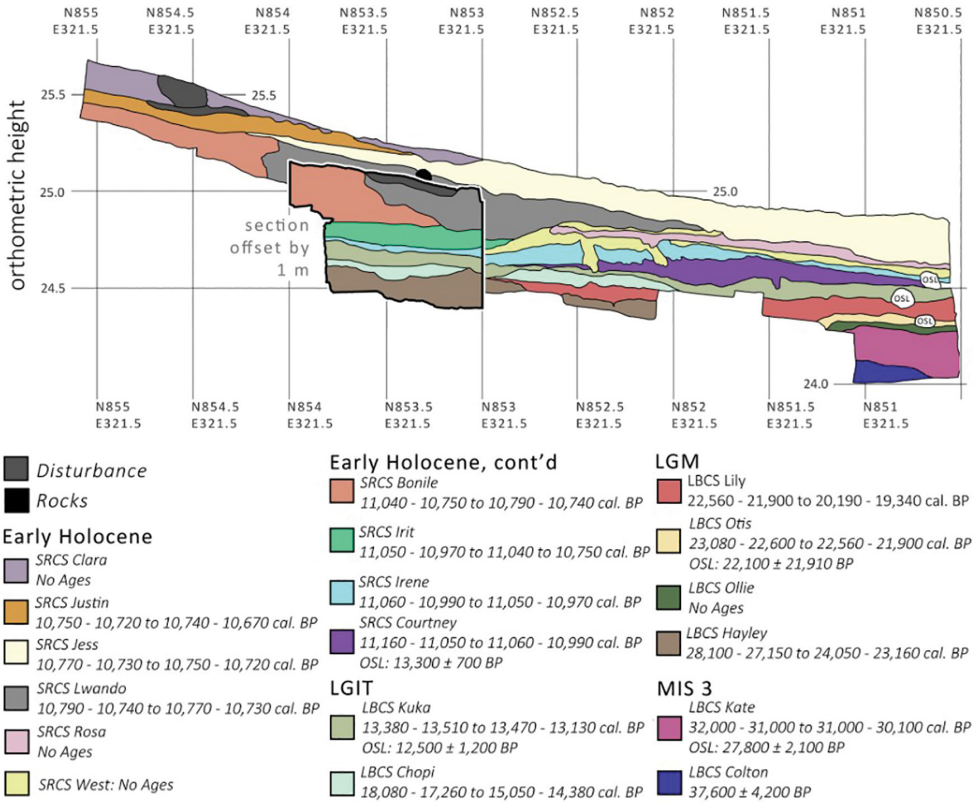
Method and materials

The fauna documented in this report was recovered from the early Holocene SRCS and the Late Pleistocene LBCS deposits during the 2016 excavations. Excavations were permitted by the Eastern Cape Provincial Heritage Resources Authority, permit #2/2/APM-PERMIT/15/03/001. The remains were analysed following Klein and

EDITORS:
Jemma Finch
Tim Forssman

KEYWORDS:
palaeoenvironment, archaeozoology, Pleistocene–Holocene transition, Marine Isotope Stage 3, Last Glacial Maximum

FUNDING:
US National Science Foundation (BCS-1827326), GENUS: DSI-NRF Centre of Excellence in Palaeoscience (86073), South African National Research Foundation (129689), University of the Witwatersrand (Carnegie DTA Grant)



Source: Adapted from Fisher et al.[1], with permission

Figure 1: The ages and locations of SubAggs at Waterfall Bluff.

Cruz-Uribe[9] and Driver[10]. Fauna was compared to comparative faunal collections at the Ditsong National Museum of Natural History, Pretoria. Bovidae (bovids) not assigned to a Linnaean Family were categorised according to four size classes based on Brain[11] from the smallest (Bov 1; e.g. oribi) through to the largest (Bov 4; e.g. eland). Specimens that could only be identified as 'mammal' but whose size could be estimated were classed as Size 1 through 4. This report includes only specimens identified to taxa, presented according to the number of identified specimens (NISP). Ungulates were categorised as grazers, browsers and mixed feeders, following Skinner and Chimimba[6]. Table 1 presents the identified fauna. A taphonomic analysis was not conducted; however, observable marks such as cut marks, gypsum or manganese staining were noted.

Results

Most specimens (NISP = 65; 89%) were from SRCS deposits; a small portion (NISP = 10; 13.7%) was from LBCS layers (Table 1). Most recovered specimens were poorly preserved, showing extensive charring and water damage, likely due to proximity to the drip line. Additionally, the assemblage was extensively fragmented, with most specimens being less than 2 cm long. Approximately 10% showed anthropogenic modifications like cut and percussion marks. None of these bones could be assigned to a species except for a charred astragalus with cut marks from the SRCS Jess deposits that belonged to a blue duiker (CN004835, Lot 221).

Bovids are the most common taxa, while rock hyrax (*Procavia capensis*) was the most identifiable mammal species (NISP = 7; 20.6%). A seal tooth (CN47208, Lot 303) was recovered from the LBCS SubAgg Lily (Table 1) (Figure 2). The crown has three distinct cusps and is unlike those of the more common fur seal (*Arctocephalus* spp.), which are smooth with less pronounced lateral cusps. The trident-shaped postcanine tooth corresponds to the earless seal (Phocidae), of which four species occur along the southern African coast.[6] Although worn, the tooth matches the prominent, triple-cusped crown of leopard seals, which are used to sieve krill. Based on our assessment, the tooth does not resemble the other three Phocidae species and is most likely that of a leopard seal. Notably, at least three other bone fragments (a vertebra and two unidentified bone specimens) recovered from the LBCS and SRCS layers may also belong to seals; this is based on their relatively distinct internal morphology.

Discussion

Late Pleistocene environment

The LBCS yields a small faunal sample with only two identified bovid species; neither are exclusive grazers (Table 2). Eland are mixed-feeders adaptable to various environmental conditions. The common duiker (obligatory browser) requires bushes and trees for shelter, shade and forage.

The presence of fish and seal remains in the LBCS layers indicate that the site was within the foraging distance of the coast[1] (Table 1). The neocoastal site's location is adjacent to the narrow continental shelf meant it was

Table 1: Number of identified specimens (NISP) from Waterfall Bluff

Taxa	Common name	SRCS	LBCS	Total
Hydrurga leptonyx	Leopard seal		1	1
Cercopithecus pygerythrus	Vervet monkey	1		1
Procavia capensis	Rock hyrax	4	3	7
Sylvicapra grimmia	Common duiker		1	1
Tragelaphus oryx	Eland		4	4
*Syncerus afer**	African buffalo	1		1
Redunca fulvorufula	Mountain reedbuck	1		1
Redunca arundinum	Southern reedbuck	1		1
Redunca sp.	Reedbuck	2		2
Philantomba monticola	Blue duiker	1		1
Ourebia ourebi	Oribi	1		1
Damaliscus pygargus	Bontebok/blesbok	1		1
Pelea capreolus	Grey rhebok	2		2
Tragelaphus scriptus	Bushbuck	4		4
Raphicerus cf. *campestris*	Steenbok	5		5
Bovidae: Indet.	Bov 1	12		12
	Bov 2	10		10
	Bov 3	2		2
	Bov 4	4		4
Bathyergidae sp.	Mole rat		1	1
Hystrix africaeaustralis	Porcupine	1		1
Testudinidae	Tortoise	7		7
Varanus sp.	Monitor lizard	2		2
Aves	Bird – small	3		3
Total		67	10	75

SRCS, Shell-Rich Clayey Sands; LBCS, Light Brown Coarse Sands

**Given the racial offensiveness of the original species name* Syncerus caffer*, we follow recent recommendations for ethical scientific naming by Smith and Figueiredo[12] and Roksandic et al.[13] by removing the c and the second f from the offensive epithet.*

Table 2: Ungulate dietary categories at Waterfall Bluff based on the number of identified specimens (NISP) and the number of taxa (NTAXA)

Dietary preference	SRCS	LBCS	Total NISP	NTAXA
Grazers	7	0	7	5
Mixed feeders	0	4	4	1
Browsers	16	4	20	6

SRCS, Shell-Rich Clayey Sands; LBCS, Light Brown Coarse Sands

within the foraging range of the coast throughout glacial times.[1] The identification of several possible barnacle remains in the sample, along with previous finds of marine shellfish, also shows that WB hunter-gatherers were systematically exploiting coastal resources even when the coast was ~8 km away.[1-3]

The leopard seal tooth (Figure 2) is from SubAgg Lily in LBCS, dated to the LGM, with a modelled age range of 22 560–19 340 cal. BP.[1] Leopard seals are apex predators whose habitats are restricted to pack ice around the Antarctic and with a maximum range to the sub-Antarctic islands.[14,15] Thus, leopard seals could have occurred further north than their usual home ranges during the peak of the LGM, which could also lead to leopard seal bodies being washed up on the coasts near WB more frequently. This is the first occurrence of a leopard seal recovered from an archaeological site off the southern African coast. The only modern records of leopard seals in South Africa were in East London in 1946, Hout Bay in 1969[6], and a dead leopard seal reported to the East London museum in 1994. More recently, Kevin Cole, Principal Scientist at the East London Museum, spotted a living individual near Cove Rock in 2014 (Figure 3) (Cole K, personal communication). Another living specimen was sighted in 2021 at Kommetjie beach in Cape Town, and the same seal was sighted again in Yzerfontein on the West Coast.[16]

The presence of a leopard seal in LBCS at WB may suggest that their colonies occurred in territories beyond the Antarctic and sub-Antarctic oceans. It is also possible that this was an isolated instance of a leopard seal straying too far from its home range, or it may be the remains of a leopard seal body that was washed up on the shore from the closer northern ranges of its habitat. However, the colder climatic conditions during the LGM, characterised by local sea surface temperatures approximately 3°C lower than current levels (as indicated by core MD96-2048[17]), may have provided suitable environments for leopard seal colonies[18] – or their hunting ranges – to have spread further north[19]. Extended periods of sea ice persistence may have facilitated increased leopard seal populations during the LBCS. Furthermore, increased Antarctic ice may have cut off specific krill distribution ranges and spawning grounds – a major source of food for leopard seals – around the east Antarctic and Antarctic Peninsula. This would have compelled the krill to expand their distributive range northwards for breeding purposes and forced leopard seals to follow their food source.[19,20] Furthermore, WB is close to a river mouth that would have washed inland nutrients into the ocean, thus maintaining a constant level of ocean nutrients that would have supported and propagated the marine animals upon which leopard seals preyed.

Early Holocene environment

In the SRCS layers, the relative prevalence of bushbuck and reedbuck suggests a riverine environment.[7] Although browsers dominate in terms of NISP (Table 2), the relative abundance of reedbuck, African buffalo and bontebok/blesbok at WB indicates a grassy environment and the likely presence of permanent water nearby.[6,7] Esteban et al.[2] found that summer rainfall increased during the Early Holocene period, promoting moister environments that the southern reedbuck would have favoured. Species showing the presence of freshwater suggests that the nearby Mlambomkulu River may have persisted from the LGM, and incisions of palaeo-rivers on the now-submerged continental shelf indicate active riverine run-off during this time.[2] Mountain reedbuck and grey rhebok indicate rocky, steep-sloped habitats[6,7] similar to the landscape found along major river drainages in the region.

Blue duiker and vervet monkey remains from the SRCS (Table 1) indicate the presence of forests or thickets, which they require for cover and food.[6] Blue duikers often exploit monkeys' feeding habits and eat the fallen foliage and fruits that the monkeys drop while scavenging in the treetops.[6] Furthermore, cut marks on the blue duiker specimen indicate that hunter-gatherers, rather than carnivores, were the primary accumulator of these small bovids at WB.

Conclusion

Our sample is very small, so we must be cautious when using the fauna to infer palaeoenvironmental changes. However, the remains of water-dependent herbivores in our sample indicate a consistent source of fresh water nearby. Furthermore, the relatively equal distribution of grazers and browsers tentatively suggests a mosaic environment of interdigitated habitats in the region throughout the Pleistocene–Holocene transition period.

The identification of a leopard seal is the first direct evidence of this species recovered from Pleistocene and Holocene archaeo-faunal assemblages

Figure 2: Leopard seal tooth (#CN47208, Lot 303).

Credit: Kevin Cole, with permission

Figure 3: Kevin Cole and a leopard seal at Cove Rock, 2014.

along the southern African coast. This and other palaeoenvironmental data may indicate a substantial ecological shift in the coastal environment during the LGM. Excavations at WB are ongoing, and future analyses of larger faunal samples would better contextualise the site's palaeoecology.

Acknowledgements

The P5 project thanks King Zanozuko Tyelovuyo Sigcau of AmaMpondo and the hospitality of the AmaMpondo people. The project further acknowledges the inestimable support of Nkosi Mthuthuzeli Mkwedini and the Lambasi AmaMpondo community for their continued interest in our research at Waterfall Bluff. P5 also thanks Lungiswa Sihlobo of the Ingquza Hill Municipality and local tourism associations for their support throughout the broader region and in promoting ecotourism at the site. We also want to acknowledge the Eastern Cape Provincial Heritage Resources Authority, the South African Heritage Resources Authority, the East London Museum, Annie Antonites and staff of the Archaeozoology Collection at the Ditsong National Museum of Natural History, and the Albany Museum of South Africa. We thank the two anonymous reviewers for their valuable comments. Their suggestions have made this a significantly better paper.

Funding

The field research was made possible through the support of a grant from the US National Science Foundation (BCS-1827326 to E.C.F., H.C.C.,

I.E. and J.P.). J.P.R. is funded by grants from GENUS: DSI-NRF Centre of Excellence in Palaeoscience under grant no. 86073, the National Research Foundation (Thuthuka grant number: 129689), and a Carnegie DTA Grant from the Faculty of Science, University of the Witwatersrand.

Data availability

The data supporting the results of this study are housed at the University of the Witwatersrand and are available upon request to the corresponding author.

Declarations

We have no competing interests to declare. We have no AI or LLM use to declare. Excavations at Waterfall Bluff were conducted under the auspices of the Eastern Cape Provincial Heritage Resources Authority, permit #2/2/APM-PERMIT/15/03/001-.

Authors' contributions

S.O.: Conceptualisation, methodology, faunal identification, data analysis, writing – initial draft. J.P.R.: Conceptualisation, methodology, faunal identification, data analysis, data curation, student supervision, writing – revisions. H.C.C.: Data collection, project leadership, funding acquisition, writing – revisions. I.E.: Data collection, project leadership, funding acquisition, writing – revisions. J.P.: Data collection, project leadership, funding acquisition, writing – revisions. E.C.F.: Data collection, project leadership, funding acquisition, writing – revisions. All authors read and approved the final manuscript.

References

1. Fisher EC, Poupart M, Pargeter J, Neumann FH, Marean CW, Jerardino A, et al. Archaeological reconnaissance for middle stone age sites along the Pondoland Coast, South Africa. J Paleontol. 2013;2013:104–137. http://doi.org/10.4207/PA.2013.ART82
2. Esteban I, Bamford MK, House A, Miller CS, Neumann FH, Schefuß E, et al. Coastal palaeoenvironments and hunter-gatherer plant-use at Waterfall Bluff rock shelter in Mpondoland (South Africa) from MIS 3 to the early Holocene. Quat Sci Rev. 2020;250, Art. #106664. https://doi.org/10.1016/j.quascirev.2020.106664
3. Oertle A, Szabó K, Gaqa S, Cawthra H, Esteban I, Pargeter J, et al. Identifying taxa from highly degraded shell micro-fragments in anthropogenic soils from Waterfall Bluff, South Africa. In: EGU General Assembly 2022. Vienna: EGU; 2022. p. EGU22-4341. https://doi.org/10.5194/egusphere-egu22-4341
4. Van Wyk AE, Smith GF. Regions of floristic endemism in southern Africa: A review with emphasis on succulents. Pretoria: Umdaus Press; 2001.
5. Rutherford MC, Mucina L, Lötter MC, Bredenkamp GJ, Smit JHL, Scott-Shaw CR, et al. Savannah biome. In: Rutherford M, Mucina L, editors. The vegetation of South Africa, Lesotho and Swaziland. Pretoria: SANBI; 2006. p. 429–529.
6. Skinner J, Chimimba C. The mammals of the southern African sub-region. Cambridge, UK: Cambridge University Press; 2005. https://doi.org/10.1017/CBO9781107340992
7. Furstenburg D. The new game rancher. Pretoria: Briza Publications; 2016.
8. Karkanas P, Brown KS, Fisher EC, Jacobs Z, Marean CW. Interpreting human behavior from depositional rates and combustion features through the study of sedimentary microfacies at site Pinnacle Point 5-6, South Africa. J Hum Evol. 2015;85:1–21. https://doi.org/10.1016/j.jhevol.2015.04.006
9. Klein RG, Cruz-Uribe K. The analysis of animal bones from archeological sites. Chicago, IL: University of Chicago Press; 1984.
10. Driver J. Manual for the description of vertebrate remains. Unpublished manual; 2005.
11. Brain CK. Some suggested procedures in the analysis of bone accumulations from southern African Quaternary sites. Ann Transvaal Mus. 1974;29(1):1–8.
12. Smith GF, Figueiredo E. Proposal to add a new Article 61.6 to permanently and retroactively eliminate epithets with the root caf[e]r- or caff[e]r- from the nomenclature of algae, fungi and plants. Taxon. 2021;70(6):1395–1396. https://doi.org/10.1002/tax.12622
13. Roksandic M, Musiba C, Radović P, Lindal J, Wu XJ, Figueiredo E, et al. Change in biological nomenclature is overdue and possible. Nat Ecol Evol. 2023;7(8):1166–1167. https://doi.org/10.1038/s41559-023-02104-x
14. Siniff DB. An overview of the ecology of Antarctic seals. Am Zool. 1991;31(1):143–149. https://doi.org/10.1093/icb/31.1.143
15. Rogers TL. Leopard seal: *Hydrurga leptonyx*. In: Perrin WF, Würsig JGM, Thewissen B. Encyclopedia of marine mammals. Cambridge, MA: Academic Press; 2009. p. 673–674. https://doi.org/10.1016/B978-0-12-373553-9.00155-3
16. Froneman A. Leopard seal seen in Cape Town makes its way to Yzerfontein [webpage on the Internet]. c2021 [cited 2023 Dec 10]. Available from: https://www.getaway.co.za/travel-news/leopard-seal-seen-in-cape-town-makes-its-way-to-yzerfontein/
17. Caley T, Kim JH, Malaizé B, Giraudeau J, Laepple T, Caillon N, et al. High-latitude obliquity as a dominant forcing in the Agulhas current system. Clim Past. 2011;7(4):1285–1296. https://doi.org/10.5194/cp-7-1285-2011
18. Testa JW, Oehlert G, Ainley DG, Bengtson JL, Siniff DB, Laws RM, et al. Temporal variability in Antarctic marine ecosystems: Periodic fluctuations in the phocid seals. Can J Fish Aquat Sci. 1991;48(4):631–639. https://doi.org/10.1139/f91-081
19. Bonillas-Monge M. Historical dynamics of leopard seal (*Hydrurga leptonyx*) and southern elephant seal (*Mirounga leonina*) populations in the Southern Ocean [unpublished doctoral dissertation]. Durham: Durham University; 2018.
20. Spiridonov VA. A scenario of the late-Pleistocene-Holocene changes in the distributional range of Antarctic krill (*Euphausia superba*). Mar Ecol. 1996;17(1–3):519–541. https://doi.org/10.1111/j.1439-0485.1996.tb00525.x

Thinking about 'service delivery'

Thirty years into South Africa's democracy, and shortly before the 2024 elections, we at the *South African Journal of Science* decided to open a call for a discussion series on service delivery, for publication after the elections early in the term of whatever new government would emerge. This series follows on our more focused Discussions on Load Shedding in 2023. It is perhaps instructive that, although at the time of writing this (early August 2024) the issue of load shedding seems less critical than a year ago, issues related more broadly to 'service delivery' are more varied and, unfortunately, durable.

When we made the call for submissions for 'Discussions on Service Delivery', we deliberately threw the net very wide and did not know how many submissions we would receive. We have been very heartened by the fact that we are publishing no fewer than 12, very varied, submissions. The first of these, in itself, makes the series worthwhile. In our call for submissions, we had not sufficiently thought about the potential difficulties with the concept of 'service delivery'. Friedman's contribution, in the best tradition of contributions to an interdisciplinary journal, forces us all to consider carefully what we mean when we talk about 'service delivery', and what is at stake politically and ideologically when we use the term. Not all readers may agree with Friedman's claim that "'service delivery' exalts the technical function of the government and diminishes the agency of the citizen", but Friedman's contribution is important in showing us that the question of 'service delivery' and its framing in this way, is in itself a political question, and not merely a technical one.

Du Plessis and Fuo raise similarly fundamental questions about the role of legislation, the courts and the judiciary in relation to provision of services, and address the thorny question of, as they put it, "where the law struggles to deliver". Here, we have a personification of the law as a potential 'deliverer', and as the authors show, it is impossible to separate questions of law from much broader and more tricky social questions. In terms of the interface between technical questions and human factors, Inggs and du Toit, although clearly aware of what they term organisational and cultural factors in how local government players organise and consider services, suggest that more attention be paid to fundamental systems analysis, with usable data forming part of feedback loops. They provide a very simple model onto which a range of contingencies can be mapped. At a more complex level, Biljohn and Magaiza argue that a quadruple helix model may go some way to help achieve, as they put it, "transformative aspirations and development futures".

In this regard, Maree and Khanyile use survey data from Gauteng to demonstrate the need for reliable and dependable infrastructure as basic for quality improvement. Given widespread concerns about water availability and quality in many South African contexts, two of the discussion pieces focus specifically on water as a key issue. Luyaba and colleagues, in their national assessment of water infrastructure, demonstrate – through their analysis of the MuSSA (Municipal Strategic Self-Assessment) framework – the complex interplay between technical, human, and social factors in questions of whether municipalities are, as they put it, unwilling or unable to deliver on promises, or both unwilling *and* unable. Focusing more closely on Johannesburg and surrounds, Sheridan emphasises the extent to which high-level questions about what we value as a society interact with more technical and visceral concerns (such as what he terms the 'yuck' factor in dealing with perceptions about drinking treated sewage), with all levels to be considered together as part of the system. Because Sheridan's contribution deals specifically with Johannesburg, we did approach Johannesburg Water for a response; at the time of going to press we had not heard back from Johannesburg Water. When considering waste issues more generally, Kalina and Schenk emphasise the importance of, as they put it, "more inclusive arrangements that involve all actors in the waste value chain, in particular communities and the informal sector", thus adding to a cross-cutting theme of social change and participation as central to dealing with what at face value may be viewed as purely technical questions.

Moving to more obviously human-centred questions, Hlongwane suggests what she terms a "varied and collaborative" approach to what is likely to be a growing challenge of long-term care for elderly South Africans, in a context of complex social change. At the other end of the age spectrum, Samuels shows what is at stake in terms of early intervention for children with disabilities; both Samuels and Hlongwane emphasise the importance of appropriate training in care work. Moving to the crucial lifespan role played by family physicians, Mash and Nash note that the holistic approach adopted by these physicians has been insufficiently considered in discussions of changes in healthcare provision, including in discussions about the implementation of the National Health Insurance system. Turning the gaze more directly on to the lived experience of healthcare providers themselves, Hoare and Mattison draw on the work of Jameton and Boss to discuss moral distress, which Hoare and Mattison describe as "a psychological and emotional response experienced by healthcare professionals when they believe they know the morally right course of action but are unable to act accordingly due to various constraints such as institutional policies, hierarchical structures, legal and ethical dilemmas, or conflicting values within a healthcare setting".

Hoare and Mattison's analysis of the question of moral distress in an overburdened healthcare system is important in itself, but it also returns the discussion series implicitly to other sectors. Friedman, as we have noted, problematises the concept of 'service delivery', rejecting a technicist, consumerist model, and we believe that in contemporary South Africa the concept of moral distress may be used across the board to describe some of the experiences of many trying to provide their best in terms of care, water, waste disposal and usage, electricity and services more generally within a less than fully enabling legal, policy, political, and social environment. No scientist or researcher can solve these challenges alone; we are grateful to all contributors to this discussion series for demonstrating the importance of thinking and working together.

HOW TO CITE:
Swartz L. Thinking about 'service delivery'. S Afr J Sci. 2024;120(11/12), Art. #19350. https://doi.org/10.17159/sajs.2024/19350

AUTHORS:
Lubabalo Luyaba[1,2]
Pilate Moyo[1]
Nonjabulo Mbhele[1]
Mukundi Mochotlhoane[1]

AFFILIATIONS:
[1]Urban and Public Infrastructure Research Initiative (UPIRI), Department of Civil Engineering, University of Cape Town, Cape Town, South Africa
[2]South African Local Government Association, Cape Town, South Africa

CORRESPONDENCE TO:
Pilate Moyo

EMAIL:
Pilate.Moyo@uct.ac.za

HOW TO CITE:
Luyaba L, Moyo P, Mbhele N, Mochotlhoane M. Unwilling or unable? A critical reflection on the state of municipal water services, 2019–2024. S Afr J Sci. 2024;120(11/12), Art. #19046. https://doi.org/10.17159/sajs.2024/19046

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
service delivery, water services authorities, municipal strategic self-assessment, MuSSA, infrastructure management

PUBLISHED:
30 August 2024

Unwilling or unable? A critical reflection on the state of municipal water services, 2019–2024

Significance:

The South African Constitution entrusts basic service provision to municipalities. Water and sanitation services are among these basic services. This paper provides a substantiated perspective on the current sub-optimal state of municipal water and sanitation services. Municipal water and sanitation services are considered for the 2019–2024 period using a seven-pillar assessment framework to evaluate whether the status quo is underlain by unwillingness or inability (or both) on the part of municipalities. The analysis shows that there is much room for improvement and identifies inefficiency as a critical priority area for improvement.

Introduction

Water and sanitation infrastructure is crucial to national health, economic development and environmental conservation. It generally forms the first line of defence against communicable diseases such as cholera and dysentery. Thus, the current government has often used the motto, "water is life, and sanitation is dignity". According to the South African Institution of Civil Engineering (SAICE), the current state of water and sanitation infrastructure in South Africa is poor.[1] This situation is expected to deteriorate with increasing urbanisation, thus putting human health and economic development at risk. While the South African government has made efforts to invest in infrastructure, the outcomes generally do not match the investment as reflected in the Auditor General's reports.[2,3] It is thus critical to have a deeper understanding of some of the underlying causes of the poor condition of South Africa's water and sanitation services to develop impactful future solutions.

In this Perspective, we examine the performance of municipalities in delivering water and sanitation services by reviewing their outcomes in seven areas as presented below. The approach provides an objective measure of assessing municipalities' performance using credible data for each of the seven areas. While the results provide the current performance levels of municipalities for water and sanitation services, the analysis can be extended to other municipal services and periods (years).

Developing a comprehensive assessment framework

To develop a comprehensive understanding of the status quo of water and sanitation services, we developed a seven-pillar framework evaluated from four perspectives. The seven pillars are: (1) infrastructure planning, (2) infrastructure delivery, (3) infrastructure operations and maintenance, (4) financial health, (5) technical capacity, (6) transversal functionality and (7) an enabling environment. The four perspectives are: municipal (administration), community (the serviced and paying), national government (regulator, enabler and supporter), and independent (objective and outside government). The perspectives (equally weighted) are then aggregated to determine an overall performance. Each perspective is graded based on three levels: good, average and poor. The gradings consider applicable indicators corresponding to the seven pillars (performance areas).

The seven pillars and their subsequent indicators are primarily based on the national Department of Water and Sanitation's Municipal Strategic Self-Assessment (MuSSA) framework for effective water services management, as shown in Figure 1. Clear and measurable indicators are required to objectively assess each of the seven pillars from the four perspectives. Table 1 presents the indicators corresponding to each performance area (and their data source). This approach was adopted as it leverages an existing government methodology that is widely used by the 144 municipalities that are water services authorities (WSAs) in South Africa.

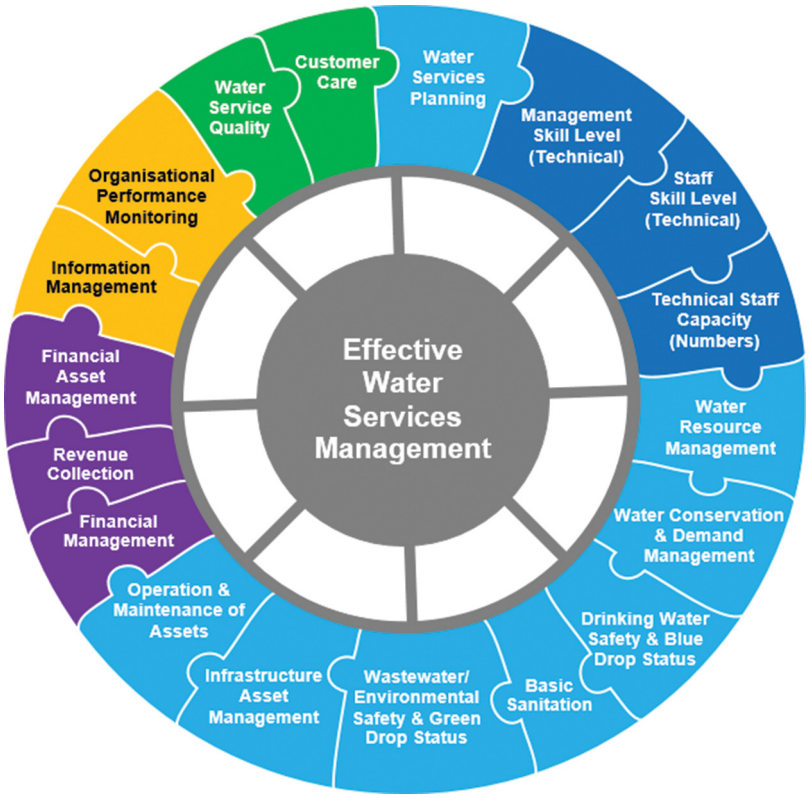
Framework results

This section presents the analysis results obtained using the seven-pillar framework, from the four perspectives: municipal, community, national government and independent.

Municipal perspective

The municipal perspective is drawn directly from the MuSSA[4], which the 144 WSAs undertake annually. The MuSSA consists of five strategic questions across the 18 areas shown in Figure 1 (note that the 18 areas constitute six of the seven pillars used in this framework). The municipalities, therefore, provide 90 (18x5) responses on their vulnerability regarding their effectiveness in providing water and sanitation services. Table 2 summarises the annual self-evaluations of the 144 WSAs from 2019 to 2023 and provides the average performance over the 5 years. Figure 2 shows a map of the 2022 MuSSA performance for each of the 144 WSAs across South Africa.

The top five (most reoccurring) challenges from 2019 to 2024 were[4]: (1) financial asset management, (2) wastewater and environmental compliance, (3) revenue collection, (4) operations and maintenance of infrastructure assets, (5) infrastructure asset management and (tied for fifth) (5) water conservation and water demand management. It



Source: © Department of Water and Sanitation[4], reproduced with permission.

Figure 1: The Municipal Strategic Self-Assessment (MuSSA) framework of the South African Department of Water and Sanitation.

must be noted that there is a lack of acceptable improvement (progression to moderate and low vulnerability) and consistent improvement in the performance of municipalities on a year-to-year basis. While some provinces moved from extreme to high and high to moderate, there was no progression towards low vulnerability. Questions should, therefore, be asked about the responsiveness and efficacy of support provided to municipalities by the national government, as it does not seem to be making a difference from this one perspective.

Community perspective

The *Local Government Municipal Systems Act*[13] outlines the roles and responsibilities of communities in a municipality. We used the Water Research Commission's (WRC) report, *The Water Services Barometer Study – User Perceptions of the Current Provision of Water Services in South Africa*[14]. The WRC study was undertaken three times – in 2011, 2015 and 2022 – and therefore provides a useful trend analysis over a reasonable period of time. The WRC studies are syndicated and undertaken on OMNIBUS and are area-stratified to be representative.

The studies focus on various issues; for this paper, we extracted the perceptions of communities about: (1) water quality, (2) reliability of water and sanitation services, (3) willingness to pay for water and sanitation services and (4) customer satisfaction. The WRC surveys showed that the urban community's perception of water quality decreased from 88% in 2015 to 78% in 2022, while the urban perception of the reliability of services decreased from 82% in 2015 to 65% in 2022.[14] Only 41% of consumers were paying for water (with only 15% knowing the exact amount).[14] Of the 59% who were not paying, 40% believed the cost was "nothing"; of this, 32% chose not to pay (up from 20% in 2015), indicating a rise in unwillingness to pay.[14] The decrease in reliability is supported by the findings of the SAICE report card[1] and the operations and maintenance challenge identified in the MuSSA[4]. The revenue collection challenges identified in the MuSSA[4] are also corroborated by the community data.

National government perspective

For the national government perspective of municipal water and sanitation, we considered the Blue Drop Report[9], the No-Drop Report[10], the Green Drop Report[11], and the Auditor General Reports[2,3]. The overall performance is a summation of the results of these four publicly available assessments.

The results were calculated as follows [*refers to the weighting of each indicator towards the overall average]:

- Blue Drop Report[9] scores (*30%), where good is 80–100%, average is 50–80% and poor is less than 50%
- No-Drop Report[10] scores (*30%), where good is a non-revenue water (NRW) of 0–20%, average is a NRW of 20–30% and poor is a NRW>30%
- Green Drop Report[11] scores (*30%), where good is 80–100%, average is 50–80% and poor is less than 50%
- Auditor General Report outcomes[2,3] (*10%), where good is 'clean and unqualified', average is 'qualified' and poor is 'adverse, disclaimer and outstanding'

Table 3 shows a national perspective consolidated from the four sources. It is clear that municipalities are struggling with wastewater management (Green Drop[11]) and water conservation and demand management (No-Drop[9]), as corroborated by the MuSSA[4] above.

Independent perspective

The 'independent perspective' relied on two complementary criteria: water and sanitation infrastructure management efficiency[15] and the ability to invest in repairs and maintenance. Efficiency refers to the relationship between inputs (available resources) and outputs (services rendered). High efficiency is achieved if adequate services are delivered within available resources.

Table 1: Overview of indicators (assessment criteria) and data sources for each performance area

Performance area (pillar)	Indicators	Data source
1. Infrastructure planning	Water and sanitation services planning	Integrated development plans (IDP)[5-8], municipal self-assessments (MuSSA)[4], Drop reports (Blue[9], Green[10] and No-Drop[11]), Census 2022[12]
	Water resource management	
	Water conservation and water demand management	
2. Infrastructure delivery	Water access levels	Division of revenue act (DORA)[6], MuSSA[4], Drop reports[9-11], Census 2022[12]
	Sanitation access levels	
	Grant expenditure performance	
3. Infrastructure operations and maintenance	Drinking water safety	Annual financial statements (AFS)[6], MuSSA[4], Drop reports[9-11], Census 2022[12]
	Wastewater / environmental compliance	
	Infrastructure asset management	
	Operations and maintenance of assets	
4. Financial health	Financial asset management	AFS[6], MuSSA[4], Drop reports[9-11], Census 2022[12], National Treasury (NT) reports[5-8]
	Revenue collection	
	Financial management	
	Auditor general opinion	
5. Technical capacity	Management skill level (technical)	Drop reports[9-11], MuSSA[4], Drop reports[9-11]
	Staff skill level (technical)	
	Technical staff capacity (numbers)	
6. Transversal functionality	Information management	AFS[6], Drop reports[9-11], MuSSA[4]
	Organisational performance	
	Water services quality	
	Customer care	
7. Enabling environment	Policy landscape	MuSSA[4], DORA[5], AG reports[2,3], NT reports[5-8]
	Regulatory landscape	
	Responsiveness and efficacy of support to municipalities (timing, quality and quantity)	

The University of Cape Town's Urban and Public Infrastructure Research Initiative (UPIRI) has developed what is called the Municipal Water and Sanitation Infrastructure Management Efficiency (MWaSSIME) Index.[15] The MWaSSIME measures how well a municipality manages its water and sanitation infrastructure compared to an ideal municipality using reported data on parameters such as water losses, water quality, and expenditure on repairs and maintenance.[15] The MWaSSIME Index utilises the Data Envelopment Analysis (DEA). The DEA was selected after considering other non-parametric and parametric methods used to measure efficiency, such as the Free Disposal Hull (FDH), the Stochastic Frontier Approach (SFA) and the Ordinary Least Squares (OLS) method.[14] The DEA assesses the efficiency of municipalities by analysing the relationships between input (resources) and output (service delivery).[16] To ensure a useful comparison, the 144 WSAs were clustered by municipal category as follows: Metros (A), Secondary Cities (B1), Large Towns (B2), Small Towns (B3), Rural Small Towns (B4) and Rural Districts (C2), as shown in Figure 3. The 144 WSAs are constituted of 8 Metros (A), 115 Locals (B) and 21 Districts (C2).

As shown in Figure 3, all South African WSAs are functioning below 50% of the ideal in regard to their efficiency in the management of water and sanitation infrastructure. The average level of infrastructure management efficiency in the Metros decreased from 33% in 2018 to 26% in 2023, and all municipalities decreased on average from 2018 to 2023. It is interesting to note that Secondary Cities consistently outperformed the Metros from 2018 to 2023. A causal link for further investigation would be the impact of political (Council) stability on infrastructure management efficiency.

The second independent analysis considers the ability to invest in infrastructure repair and maintenance. One of the root causes of poor service delivery is underinvestment in existing infrastructure, as evidenced by low expenditure on repair and maintenance, which generally results from:

a) a systemic shortage of funding in certain municipalities (an inability to invest in operations and maintenance); and/ or

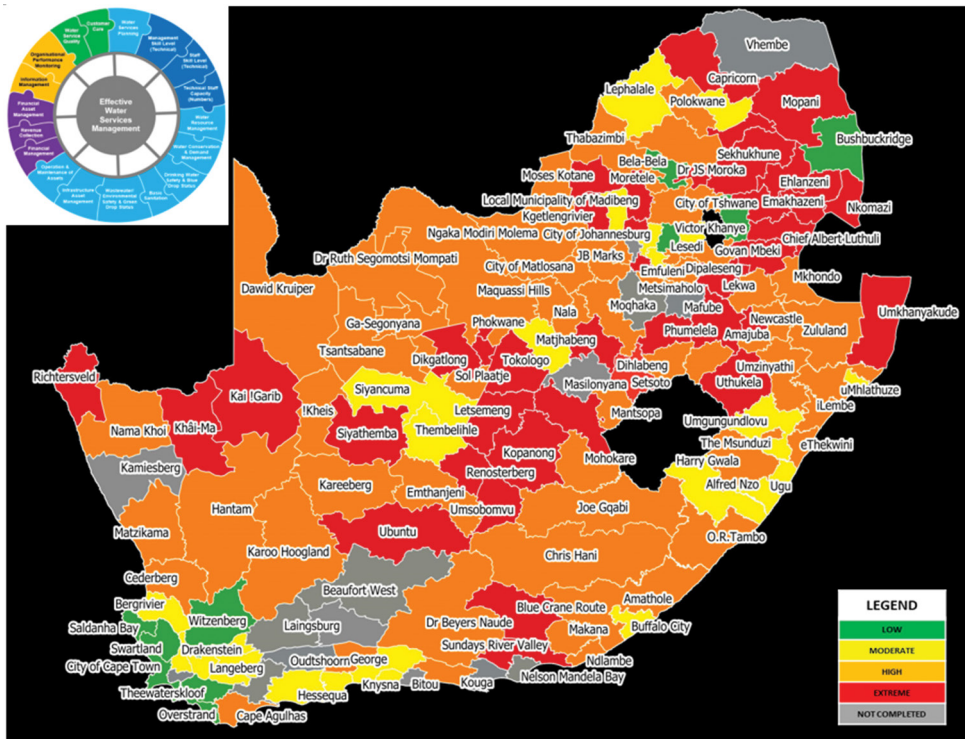
b) a failure (unwillingness) to spend on existing infrastructure, even when some funding is available.

In the first case (inability), municipalities, despite their best efforts, cannot generate enough money to invest in the upkeep of existing infrastructure. This is a municipal finance question as it speaks to municipal budgets,

Table 2: A 5-year overview of municipal vulnerability from a municipal perspective[4]

	Municipal self-assessed vulnerability per year					
Province	**2019**	**2020**	**2021**	**2022**	**2023**	**Average**
Eastern Cape						High
Free State						Extreme
Gauteng						High
KwaZulu-Natal						High
Limpopo						High
Mpumalanga						High
Northern Cape						Extreme
North West						High
Western Cape						Moderate
Average	High	High	High	High	High	High

VULNERABILITY KEY			
EXTREME	HIGH	MODERATE	LOW



Source: © Department of Water and Sanitation[4], reproduced with permission.

Figure 2: A national overview of the 2022 MuSSA results.[4]

expenditure and financial viability, and ultimately raises questions on the local government equitable share and the division of revenue and, most fundamentally, the assumptions that underpin the funding model to local government; all of which are the purview of National Treasury.

To determine which municipalities do not have enough financial resources to maintain existing infrastructure, we use a simplified formula that compares total available revenue to the amount needed to repair and maintain existing infrastructure effectively. The assumption is that the repairs and maintenance should not exceed 25% of a municipality's revenue and that 8% of the value of property, plant and equipment (PPE) should be set aside for repairs and maintenance.[5] This translates to the criterion shown by Equation 1:

$$\frac{8\%\ of\ PPE}{Total\ Potential\ Revenue} \leq 25\% \qquad \text{Equation 1}$$

Table 3: Summary overview of municipal performance from a national perspective[2,9-11]

Province	Green Drop	Blue Drop	No Drop	Auditor General outcome	Average
Eastern Cape	Poor	Average	Poor	Average	Poor
Free State	Poor	Poor	Poor	Poor	Poor
Gauteng	Average	Good	Poor	Good	Average
KwaZulu-Natal	Poor	Average	Poor	Good	Average
Limpopo	Poor	Average	Poor	Average	Poor
Mpumalanga	Poor	Average	Poor	Average	Poor
Northern Cape	Poor	Average	Poor	Average	Poor
North West	Poor	Poor	Poor	Poor	Poor
Western Cape	Average	Good	Average	Good	Average
Average	Poor	Average	Poor	Average	Poor

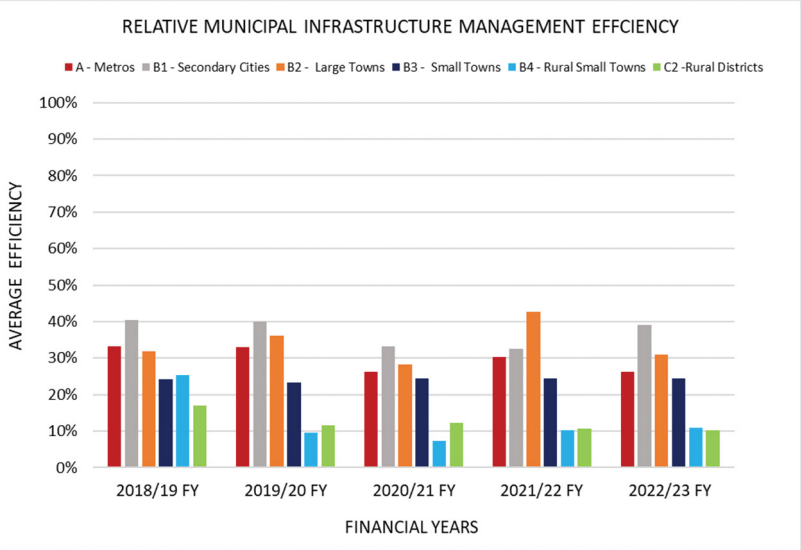


Figure 3: Relative infrastructure management efficiency of water services authorities over a 5-year period, 2018/2019 to 2022/2023.

Model assumptions:

a) Audited data provided in the municipal annual financial statements are accurate and true as legally expected in the *Municipal Finance Management Act* (MFMA).[17]

b) The 8% requirement for repairs and maintenance as per National Treasury Circular 71[7] of 2014 is the minimum amount required to maintain existing infrastructure effectively.

c) All existing infrastructure is ageing and included in PPE. The MFMA (Section 63) enjoins municipalities to maintain all the infrastructure (assets).

d) Anything above 25% of total revenue is unreasonable as municipalities have other functions to perform.

By applying the criterion, the following results are obtained: 40% (58 of 144) of all WSAs are identified as unable to maintain existing assets, with 71% (15 of the 21) of Rural Districts (C2) appearing on this list, 44% (4 of 9) of Rural Small Towns (B4) and 40% (27 of 68) of Small Towns (B3). This finding supports the view that a lack of financial resources makes the efficient management of water and sanitation infrastructure extremely unlikely, as it takes revenue (equitable share and own revenue) to look after existing infrastructure. This again points to an environment that is not enabling, especially for more grant-reliant municipalities. This group of municipalities must be treated differently when considering support and solutions.

A summary overview of performance

The Constitution recognises the three spheres of government as distinctive, interdependent and interrelated.[18] The three spheres are expected to work as one in delivering services. Table 4 summarises how the government has performed on the Medium-Term Strategic Framework (MTSF) for 2019–2024.[19] The government has set these targets, and performance is generally poor.

While the performance is poor, one can still appreciate that the targets were appropriate. The progress made in the sub-activities indicates efforts made by the sector department (the national Department of Water and Sanitation (DWS), and time may be required to see if the positive momentum translates into tangible and lasting change.

Table 5 summarises the state of municipal water and sanitation services across the seven pillars in the 2019–2024 period. While the picture looks very dire, there is ample opportunity for change. What will be key is selecting the appropriate priorities and moving as one. From the analysis above, it should be abundantly clear that there is no silver bullet for municipal water and sanitation service delivery.

Table 4: Government's Medium-Term Strategic Framework (2019–2024) target achievement[19]

No	Target	Achieved (Yes/No)	Activities	Achieved
1	100% of water services authorities (WSAs) have acceptable MuSSA scores	No	Annual assessment of all WSAs	No
			WSAs being supported to develop and implement	Partial
2	90% have access to sanitation and hygiene	No	Development and implementation of the National Sanitation Integrated Plan	Partial
			Eradication of the bucket system	No
3	95% reliability of water services	No	Refurbishment projects to address functionality of reliability implementation plans	Partial
			Blue Drop assessment and compliance	Partial
			Non-compliance monitoring	Yes
			District Municipal 5-year reliability plans	Partial
4	100% wastewater treatment works functionality	No	Bulk projects implementation	Partial
			Green Drop assessments and compliance	Partial
			Wastewater system monitoring against regulatory standards	Yes

Table 5: A summary overview of performance from 2019 to 2024

Performance area	Municipal perspective	Community perspective	National government perspective	Independent perspective	Overall performance
Infrastructure planning	Good	Average	Average	Average	Average
Infrastructure delivery	Good	Average	Average	Poor	Average
Infrastructure operations and maintenance	Poor	Poor	Poor	Poor	Poor
Financial health	Poor	Good	Average	Poor	Average
Technical capacity	Poor	Poor	Poor	Average	Poor
Transversal functionality	Average	Poor	Poor	Average	Poor
Enabling environment	Poor	Poor	Poor	Poor	Poor

Conclusions and recommendations

Due to the impact of water and sanitation services on societal well-being, economic development and the natural environment, these services must be provided efficiently and effectively. The results from the analysis suggest that a lot more work is required if South Africa is to say that the majority of the population receives reliable and sustainable basic services. The high-level conclusions and recommendations for each performance area (pillar) are:

Infrastructure planning – Mechanisms to measure the efficacy of municipal water and sanitation infrastructure planning are required. This area is not meaningfully assessed through any intergovernmental mechanisms, even though we all recognise the importance of infrastructure planning.

Infrastructure delivery – Inefficiency in infrastructure delivery needs to be urgently addressed; between 2011 and 2022, access to water only increased by 0.1% (91.2% to 91.3%) and sanitation by 11.5% (71.3% to 82.3%), despite an investment of nearly ZAR20 billion per annum into municipal water and sanitation.[5,12] The government should not only focus on expenditure monitoring for conditional infrastructure grants but also implement a more holistic approach, as Ndalasi et al.[20] argued.

Infrastructure operations and maintenance – This is a neglected area that needs urgent attention from both a financial and skills perspective. The funding model and the extent to which it allows municipalities to balance between CAPEX (capital expenditure) and OPEX (operational expenditure) should also be closely examined.

Financial health – This area is complex, as there are problems with the funding model, municipal financial management[2], and end-user willingness to pay. These must be attended to simultaneously, but more money alone is not the answer.

Technical capacity – Much has been said and written about skills in municipalities, but these commentaries seldom appreciate that the problem is three-pronged: attracting, retaining and affording. Alternative models must be explored to address the problem on all three fronts.

Transversal functionality – While seen as 'softer' issues, information management, organisational performance and customer care require attention to improve the quality of water services. Customer knowledge of their obligations also warrants a closer investigation across all Living Standard Measures (LSMs).[17]

Enabling environment – Care must be taken to strengthen the level and quality of monitoring and evaluation of provincial and national governments, as their shortcomings inevitably affect municipalities. More accountability for supporting and strengthening the capacity of municipalities is required. The MuSSA results[4] suggest that the quality of support provided to municipalities is ineffective.

From the analysis, we conclude that municipalities are both unwilling and unable to deliver quality water and sanitation services. On the one hand, their administration and infrastructure management is poor; on the other hand, they are constrained by inadequate support from the other spheres of government and a critical shortage of funds (for services rendered and from the fiscus). We argue that existing legislative and administrative processes can address some of these challenges.

Acknowledgements

Preliminary data collection and cleaning from students in the Data for Municipal Infrastructure Assets (#DATA4MIA) vacation work programme are acknowledged. #DATA4MIA programme support from UPIRI, the Municipal Infrastructure Support Agent (MISA) and SALGA is acknowledged.

Data availability

The data supporting the results of this study are available upon request to the corresponding author.

Declarations

We have no competing interests to declare. We have no AI or LLM use to declare. All authors read and approved the final version.

References

1. South African Institution of Civil Engineering (SAICE). Infrastructure report card for South Africa. Johannesburg: SAICE; 2022 [cited 2024 May 17]. Available from: https://saice.org.za/downloads/SAICE-2022-Infrastructure-Report-Card.pdf
2. Auditor General South Africa (AGSA). Consolidated general report on local government outcomes: MFMA 2021-22. Pretoria: AGSA; 2023 [cited 2024 May 17]. Available from: https://www.agsa.co.za/Reporting/MFMAReports.aspx
3. Auditor General South Africa (AGSA). Consolidated general report on local government outcomes: MFMA 2018-19. Pretoria: AGSA; 2020 [cited 2024 May 17]. Available from: https://www.agsa.co.za/Reporting/MFMAReports.aspx
4. South African Department of Water and Sanitation. Municipal strategic self assessments [webpage on the Internet]. c2024 [cited 2024 May 17]. Available from: https://ws.dws.gov.za/mussa/Default.aspx
5. National Treasury of South Africa. Division of revenue acts 2004 to 2023 [webpage on the Internet]. c2023 [cited 2024 May 17]. Available from: https://www.treasury.gov.za/legislation/bills/default.aspx
6. National Treasury of South Africa. Review of local government infrastructure grants – recommendations for reform. Draft report to Budget Forum [document on the Internet]. c2014 [cited 2024 May 17]. Available from: https://mfma.treasury.gov.za/Media_Releases/ReviewOfLGInfrastructureGrants/Documents/Reviewof LG Infrastructure Grants -% Recommendations Working Paper.pdf
7. National Treasury of South Africa. MFMA Circular no. 71 – uniform financial ratios and norms. Pretoria: National Treasury; 2014 [cited 2024 May 17]. Available from: https://mfma.treasury.gov.za/Circulars/Pages/Circular71.aspx
8. National Treasury of South Africa. Audited municipal annual financial statements. Pretoria: National Treasury; 2022.
9. South African Department of Water and Sanitation (DWS). Blue Drop report. Pretoria: DWS; 2023 [cited 2024 may 2024]. Available from: https://ws.dws.gov.za/IRIS/releases/BDWR.pdf
10. South African Department of Water and Sanitation (DWS). No-drop report. Pretoria: DWS; 2023 [cited 2024 May 17]. Available from: https://ws.dws.gov.za/IRIS/releases/ND_2023_Report.pdf
11. South African Department of Water and Sanitation (DWS). Green Drop report. Pretoria: DWS; 2022 [cited 2024 May 17]. Available from: https://ws.dws.gov.za/iris/releases/Report_DPW_Rev02_29Mar22_MN%20web.pdf
12. Statistics South Africa. Census 2022: Statistical release P0301.4. Pretoria: StatsSA; 2023 [cited on 2024 May 17]. Available from: https://census.statssa.gov.za/assets/documents/2022/P03014_Census_2022_Statistical_Release.pdf
13. Republic of South Africa. Local Municipal Systems Act, Act 32 of 2000. Pretoria: Government Printers; 2000 [cited 2024 May 17]. Available from: https://www.gov.za/documents/local-government-municipal-systems-act
14. Slabbert S. The Water Services Barometer Study 2022. User perceptions of the current provision of water services in South Africa. Report TT909-22. Pretoria: WRC; 2023 [cited 2024 May 17]. Available from: https://www.wrc.org.za/wp-content/uploads/mdocs/TT%20909.pdf
15. Mbhele N. The development of the municipal water and sanitation services infrastructure management efficiency index (using the Data Envelopment Analysis technique) [MSc thesis]. Cape Town: University of Cape Town; 2024.
16. Ramanathan R. An introduction to data envelopment analysis: A tool for performance measurement. New Delhi: Sage; 2003.
17. Republic of South Africa. Local Government: Municipal Finance Management Act 56 of 2003 [document on the Internet]. c 2004 [cited 2024 May 17]. Available from: https://www.gov.za/documents/local-government-municipal-finance-management-act-0
18. Republic of South Africa. The Constitution of the Republic of South Africa, Act 108. Pretoria: Government Printers; 1996 [cited 2024 May 17]. Available from: https://www.gov.za/documents/constitution/constitution-republic-south-africa-04-feb-1997
19. South African Department of Planning. Monitoring and evaluation. Revised Medium-Term Strategic Framework: 2019 to 2024. Pretoria: Department of Planning; 2021 [cited on 2024 May 17]. Available from: https://www.dpme.gov.za/keyfocusareas/outcomesSite/MTSF_2019_2024/Final Revised MTSF %202019-2024 2021.pdf
20. Ndalasi LA, Luyaba L, Vhele AI. Revising the municipal infrastructure grant to improve expenditure and quality outcomes. In: Proceedings of the 86th Annual IMESA Conference; 2023 October 25–27; Gqeberha, South Africa. p. 149–152. Available from: https://www.imesa.org.za/wp-content/uploads/conf2023-15.pdf

AUTHOR:
Steven Friedman[1]

AFFILIATION:
[1]Humanities Faculty, University of Johannesburg, Johannesburg, South Africa

CORRESPONDENCE TO:
Steven Friedman

EMAIL:
SFriedman@uj.ac.za

HOW TO CITE:
Friedman S. Who will free us from 'service delivery'? Why a supposed solution is a problem. S Afr J Sci. 2024;120(11/12), Art. #19326. https://doi.org/10.17159/sajs.2024/19326

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
democracy, public services, governance, citizenship, voice

PUBLISHED:
30 August 2024

Who will free us from 'service delivery'? Why a supposed solution is a problem

Significance:

In this Commentary, I challenge the conventional wisdom which sees 'service delivery' as the core task of democratic government. I argue that this view has devalued the role of citizens and prompted widespread dissatisfaction with government. The Commentary therefore points towards a fundamental rethink of government's purpose, which, if adopted, should generate new approaches which place the citizen at the heart of governance.

Few phrases are as popular when South Africans discuss government as 'service delivery'. It is assumed to be the prime purpose of governing. Its lack is said to be behind every township protest. There is some debate on how to achieve it, but none about whether to strive for it.

There are understandable reasons for this. A core feature of the apartheid system was gross inequality in the provision of public services. The suburbs in which whites lived were well served, while segregated black townships were denied basic services, initially because most black people were meant to leave the cities for rural areas – this lack of basic urban services was a core cause of the urban rebellion against apartheid. It also forced many people to rely on unauthorised services, which posed problems for the effectiveness of the state after democracy was achieved. It follows from this that a democratic government should be expected to provide the services to the majority which apartheid did not.

It is also widely agreed that South Africans, particularly the majority who apartheid dominated, do not receive the public service to which citizens of a democracy are entitled. This has prompted a distrust in the government's willingness and ability to serve the people, which is expressed in the public debate as a demand for 'service delivery'.

But 'service delivery' is not a self-evident good. On the contrary, the phrase and the thinking which produced it express a deeply undemocratic view of the relationship between the government and citizens. While its frequent use is understandable, it also makes it more difficult to achieve development goals and is the cause of much citizen dissatisfaction with government, particularly its local variety.

Degrading citizenship

The core of the democratic idea is that the people are meant to govern. This is justified because democratic thought holds that all adults are able to decide what they need and so no person, regardless of their presumed expertise, has a greater right to decide on issues affecting them than any other.[1] Because the people cannot govern directly, they establish governments. But the government is the servant of the people.

This principle is denied by 'service delivery'. It portrays the government as a 'deliverer' and the people as receivers, much like consumers in the marketplace. The people can complain when what they need is not 'delivered'. But, like consumers, they do not decide what is delivered and how it is prepared for delivery. Citizens who are relegated to recipients of delivery are worse off than consumers as they cannot choose to opt out of 'delivery' or switch to another 'deliverer' – they cannot choose not to be governed.

In 'service delivery', the government does not serve the people – it provides for them that which they are assumed to need. Citizens are not decision-makers – they are recipients of the decisions of others. They do not act but are acted upon by the 'deliverers' of service. 'Delivery' is a technical task, not the product of a relationship founded on listening and responding. It places the government over the citizenry rather than at its service. And so the concern for 'delivery' inevitably assumes a style of government which currently prevails in much of the ostensibly democratic world – one in which relatively small groups of technically qualified people decide for everyone else what the government should offer citizens and how it should offer it. If the task is purely technical, it requires people with the right training, not those who listen to citizens. Listening is an obstacle to 'delivery' because it gives credence to the view of people considered unqualified to decide what is needed.

'Service delivery' imposes a new minority rule in which the few decide for the many. At best, it reflects the 'democratic elitist' view[2] that there is a contradiction between democratic government and citizen participation[3]. Because the people do not know what the government should do and how it should do it, this view holds; democracy will collapse if the people decide because elected governments will be unable to perform the tasks required of them. Adherents of this view insist that they are democrats because they believe that the people should choose who decides for them. But they insist that this very limited right to choose will not endure if citizens insist on taking government decisions.

This view is far more a prejudice than a theory of government. There is no 'objective' measure of which policies are best for society – choices are products of our interests and values and so are always subjective. People know what their interests are and the Nobel laureate's view of what is right and good is not intrinsically better than that of the office cleaner. Unlike democratic thought, this view assumes that some of us are better than others, and it is the view which underpins 'service delivery'.

The democratic view sees the government's role as 'public service', not 'delivery'. This assumes that it is the government's function to serve citizens. The people are no longer recipients, they are the source of authority and so of decision-making. A government which serves, rather than delivers, recognises that it is the citizenry which is in charge – officials, elected or not, become what democratic principle insists that they are, public servants.

To insist on 'public service' is not to deny that governments need technical knowledge to perform their functions. Nor does it make the mistake of influential development approaches which expect citizens to choose the techniques needed to serve them.[4] Few of us are experts in public finance or the technical requirements of bulk infrastructure. Even fewer have the time or inclination to decide exactly how government technicians should meet our needs. But democracy requires that technical skills are used to implement democratic decisions, not to substitute for them. The people, through their representatives, decide what government should do and the technicians use their knowledge to do it.

The 'service delivery' approach is elitist, not because it assumes that technical knowledge is needed, but because it requires that those who possess it decide for society, rather than placing their knowledge at its service. 'Service delivery' downgrades elected representatives from decision-makers reflecting the will of the people to, at best, monitors of whether and how technicians 'deliver' services.

Imposing needs

Some imbued with zeal for 'service delivery' would dismiss this argument as theorising at the citizenry's expense. Even if 'service delivery' falls foul of democratic principle, why complain if it meets needs? But it does not meet needs – it is as flawed in practice as it is in theory.

Repeatedly since 1994, development efforts in South Africa have been limited by a gulf between what the government initiates – inevitably on the advice of 'service delivery' technicians – and what citizens who are meant to benefit want.[5] Because they do not listen to the citizens who they hope to benefit, the 'deliverers' often misread what the recipients of 'delivery' want and so their plans remain unrealisable.

The many examples will not be detailed in full here. But in the early 1990s, a core preoccupation of housing policy technicians, expressed in the work of the National Housing Forum, was how to ensure access to mortgage finance for residents of townships and shack settlements.[6] But the intended beneficiaries wanted nothing to do with mortgages, which they associated with losing their homes because they could not keep up with payments.[7] So, much time and effort were wasted negotiating how to 'deliver' to citizens what they did not want. At about the same time, mass electrification efforts assumed that people who had been denied electricity would, if offered it, replace all their sources of heating and light with electrical alternatives. They did not – research later showed that the assumption was based purely on abstract reasoning which simply did not inquire into the concrete choices facing householders.[8]

In both cases, 'service delivery' meant implementing the abstract reasoning of technicians, even though it neither reflected the choices nor suited the circumstances of those who were meant to benefit. This was a consequence of relying on technical knowledge and insights rather than hearing and responding to the voices of the citizenry. It made 'service delivery' a failure on its own terms – it failed to ensure the efficient programmes which technical expertise purports to offer because society's needs can only be established by listening carefully to the diverse voices of the people and allowing a democratic contest between ideas.

The housing example is also important because the forum was not only a gathering of development technicians, it also brought together a broad range of interests and was widely praised precisely because it was seen to be inclusive and broadly representative. That this forum should focus on a 'solution' based on fashionable theorising rather than an accurate reading of citizen need showed how deeply embedded the 'service delivery' view is among the country's elites. It remains so today.

Since the early 1990s, development policy has repeatedly relied on the assumptions of technicians, not the expressed needs and wishes of citizens. This is why research shows that by far the most effective anti-poverty measure since 1994 has been social grants[9] which have stimulated local economies and are used by recipients to generate local economic activity[10]. They are often stigmatised as sources of 'dependency' but are in reality sources of popular empowerment because they enable people to make choices. One reason for the stigmatisation is that grants contradict the assumptions of 'service delivery' because they enable citizens who receive them to make their own choices rather than living with the consequences of the decisions of 'service deliverers'.

Imposing views

'Service delivery' also muzzles the voices of most citizens.

It was noted earlier that township demonstrations are repeatedly labelled 'service delivery protests' because it is assumed that participants are demanding 'delivery'. But it is more accurate to see the demonstrations as protests against 'service delivery'.

Portraying demonstrations as demands for 'delivery' often silences participants, although it purports to give them voice. Simply labelling an event in this way absolves the journalists, commentators and politicians who use it of the bother of finding out what people really want. It can also distort what they want by filtering it through the lens of 'service delivery' when there are more accurate ways of understanding their concerns.

Citizens who complain about a lack of electricity or water are not necessarily demanding 'service delivery'. They are not necessarily insisting that the government deploy its technical skills to impose solutions on them. They might be demanding that the authorities listen carefully to their needs.

Evidence supports this view. Studies of why people protest present a far more complicated picture than the blanket 'service delivery' explanation.[11] In some cases, it could be argued that protesters are demanding not that they receive 'service delivery' but that they be freed from it. In one, technicians decided that 'service delivery' would be more effective if some residents of a shack settlement were removed from their homes. The affected householders became understandably angry and took to the streets in protest.[12] 'Service delivery' threatened their right to shelter and so they wanted to escape it. Restrictions imposed on small traders, so that they do not trade in places where their presence interferes with the intentions of planners, can also become a cause for protest[13]; here too, the reason for demonstrations is not a desire for 'service delivery' but to escape it.

Even if we were to assume, wrongly, that everyone who wants better services really wants 'service delivery', not all protests are a demand for improved services. If the media and commentators need a blanket explanation for why people take to the streets, the more credible phrase would be that these are protests by people who believe they are not heard. And they will not be heard as long as 'service delivery' takes precedence over public service.

Assuming that they want 'delivery' silences people in townships by assuming that they share the world view of the reporter or commentator using the phrase. People may also be silenced even as they seem to be heard. The media may appear to give voice to protestors or to aggrieved citizens. But the questions posed to demonstrators ensure that they offer only a litany of grievances about inadequate services. There is no attempt to learn why particular protests happen at particular places at particular times, even less to discern the choices protestors would make if they did enjoy a say in how the places in which they live would be served. The litany neatly reinforces suburban views of government's role and ensures that the few are heard through the voices of the many.

When middle-class citizens protest (very occasionally by taking to the streets but far more often by signing petitions or lobbying in other ways) their protests are not dismissed by a catch-all phrase – their grievances are described[14] and so they are heard. 'Service delivery protestors' are not. Trade unions, despite strong biases against them[15], are also likely to be heard in a way which is denied to citizens whose concerns are dissolved in the blanket phrase 'service delivery.'

The wrong cure

If 'service delivery' is what people expect of democratic government, the solution lies in more technically efficient government institutions.

'Service delivery' is so popular a term because this is precisely how governing is seen in South Africa and many other countries. Government's failure to win the approval of citizens is said to stem from its lack of capacity[16], which is always explicitly or implicitly understood as technical capacity. The 'solution' is, therefore, to enhance technical capability. This view is popular, not only among journalists and commentators, but also among aid donors and the non-governmental organisations they support.[17]

It may also be the preferred approach of the government. Some of this thinking underpins its District Development Model which assumes that efficiency will be enhanced if local, provincial, and national government tackle key tasks jointly. While the language which justifies this model talks of the benefits each sphere of government can derive from working with the others[18], it has been justified as a means of rectifying the weaknesses of local government in particular. The unstated assumption is that municipalities can improve their ability to perform technical tasks by learning from the other spheres, whose technical competence is assumed to be superior.

But, if we acknowledge that the problem is not technical capacity but a refusal to listen, enhancing 'service delivery' will not improve the way people are governed. Because 'service delivery' exalts the technical function of the government and diminishes the agency of the citizen, and citizens hold local government in low esteem not because it fails to 'deliver' to them, but because it will not listen to them, the solution lies in adopting a 'public service' approach in which the citizen is the thinking, acting agent and government is its servant, whose role is to turn what most citizens want into reality. The solution would then lie not in enhancing the government's technical capacity but in insisting that it discharge its democratic responsibilities to the people by listening to them and implementing the preferences of the majority on any particular issue.

It is also worth noting that knowing what is technically efficient is not as straightforward as advocates of 'service delivery' assume it to be. 'Efficiency' is a measure of capacity to realise goals. But who decides what the goals are? The question is political, not technical – what citizens want from government depends on their interests and values and there is no 'objective' way of determining who is 'right' and who is 'wrong'. And so, citizens are likely to be satisfied with governments which are pursuing goals which they share in a manner which responds to their needs. In the effort to achieve this, technical efficacy is secondary, willingness to listen and respond to citizen voice primary. Technical skills enable what is heard to be implemented but they are not a substitute for responding to citizens.

The approach most consistent with democratic principles is also that which is most likely to address the problem 'service delivery' is meant to solve – citizen alienation from government. A commitment to 'public service' rather than 'service delivery' is most likely to convince most citizens that government exists to serve them and so to mend the broken link, in just about all democracies including South Africa, between citizens and those who purport to serve them.

Declarations

I have no competing interests to declare. I have no AI or LLM use to declare.

References

1. Friedman S. Power in action: Democracy, citizenship and social justice. Johannesburg: Wits University Press; 2018. https://doi.org/10.18772/12018113023
2. Schumpeter J.Capitalism, socialism and democracy. London: Unwin; 1942.
3. Avritzer L. Democracy and the public space in Latin America. Princeton, NJ: Princeton University Press; 2002. p. 11.
4. Olaleye YL. The doctrine of citizens' participation in organisation and implementation of community development projects. Afr J Educ Res. 2012;16:28–34.
5. Friedman S. South Africa: Globalization and the politics of redistribution. In: Tulchin JS, Bland G, editors. Getting globalization right: The dilemmas of inequality. London: Lynne Rienner; 2005. p. 11–49. https://doi.org/10.1515/9781685851309-004
6. Huchzermeyer M. Housing for the poor? Negotiated housing policy in South Africa. Habitat Int. 2001;25:303–331. https://doi.org/10.1016/S0197-3975(00)00037-0
7. Tomlinson MR. From rejection to resignation: Beneficiaries' views on the South African government's new housing subsidy scheme. Urban Stud. 1999;36(8):1349–1359. https://doi.org/10.1080/0042098993024
8. White C, et al. The social determinants of energy use: Synthesis report. Unpublished report submitted to the Department of Mineral and Energy Affairs; January 2000; Friedman S. South Africa: Globalization and the politics of redistribution. In: Tulchin JS, Bland G, editors. Getting globalization right: The dilemmas of inequality. London: Lynne Rienner; 2005. p. 11–49. https://doi.org/10.1515/9781685851309-004
9. Van der Berg A, Krige S, Lekezwa B. Efficiency and equity effects of social grants in South Africa. Stellenbosch Economic Working Papers 15/10SSRN: 2010. https://doi.org/10.2139/ssrn.1727643
10. Seleoane M. Resource flows in poor communities: A reflection on four case studies. In: Habib A, Maharaj B, editors. Giving and solidarity: Resource flows for poverty alleviation in South Africa. Pretoria: HSRC Press; 2007. p. 121–158.
11. Von Holdt K, Langa M, Molapo S, Mogapi N, Ngubeni K, Dlamini J. The smoke that calls: Insurgent citizenship, collective violence and the search for a place in the New South Africa. Johannesburg: Centre for the Study of Violence and Reconciliation, Society, Work and Development Institute; 2011.
12. Mkhize T. Forced-removal rumours spark violent protests in Diepsloot. TimesLive. 2007 November 02. [cited 2016 Jul 12]. Available from: https://www.timeslive.co.za/news/south-africa/2007-11-02-forced-removal-rumours-spark-violent-protests-in-diepsloot/
13. Friedman S, Hlela K, Thulare P. A question of voice: Informality and pro-poor policy in Johannesburg, South Africa. In: Hamdi N, editor. Urban futures: Economic growth and poverty reduction. Rugby, UK: ITDG; 2005. p. 51–68. https://doi.org/10.3362/9781780446325.006
14. Engel K. Suburbs being "turned into slums" in Cape Town as City turns blind eye to by-law infringements. Cape Argus. 2023 May 10 [cited 2024 Mar 05]. Available from: https://www.iol.co.za/capeargus/news/suburbs-being-tuned-into-slums-into-cape-town-as-city-turns-blind-eye-to-by-law-infringements-31f89890-7dc0-44bc-b2e8-ba6b279ec0ae
15. Vabaza S. Labour unions: Why they are not good for South Africa. City Press. 2022 December 05. [cited 2024 Mar 05]. Available from: https://www.news24.com/citypress/voices/labour-unions-why-they-are-not-good-for-south-africa-20221205
16. Brand S. Local government in South Africa is in crisis. How it can be fixed. The Conversation. 2018 May 29. [cited 2022 Aug 12]. Available from: https://theconversation.com/local-government-in-south-africa-is-in-crisis-how-it-can-be-fixed-97331
17. Bloch G, Favis M, Hargovan J. Evaluation of ODA to Capacity Building. Pretoria: National Treasury; 2000. Available from: https://www.treasury.gov.za/publications/other/devco-op/section_2/01.pdf
18. South African Department of Co-Operative Government and Traditional Affairs. District development model: To accelerate, align and integrate service delivery [webpage on the Internet]. No date. [cited 2024 Mar 06]. Available from: https://www.cogta.gov.za/ddm/

AUTHORS:
Anél du Plessis[1]
Oliver N. Fuo[2]

AFFILIATIONS:
[1]Professor of Law and Chair: Urban Law and Sustainability Governance, Stellenbosch University, Stellenbosch, South Africa
[2]Professor of Law and SARChI Interim Chair: Cities, Law and Environmental Sustainability, North West University, Potchefstroom, South Africa

CORRESPONDENCE TO:
Anél du Plessis

EMAIL:
adup@sun.ac.za

HOW TO CITE:
du Plessis A, Fuo ON. When law struggles to deliver: Reflections on service delivery law reform in South Africa, 1996–2024. S Afr J Sci. 2024;120(11/12), Art. #19336. https://doi.org/10.17159/sajs.2024/19336

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
local government law, service delivery, developmental local government, law reform, accountability

PUBLISHED:
30 August 2024

When law struggles to deliver: Reflections on service delivery law reform in South Africa, 1996–2024

Significance:

This contribution reflects on service delivery law reform in post-apartheid South Africa, focusing on the expectations that were cemented in law for local government. This reflection is offered in the context of the current widespread service delivery challenges and the absence of consequence management in local government. We comment on the possible reasons why the legal framework struggles to deliver on the original policy vision of 'developmental local government' which was envisaged in the late 1990s to deliver fair, sustainable and equal future access to services of a high quality.

Introduction

In line with the principle of subsidiarity[1], it usually happens that the law or legal processes put municipalities (local authorities) in charge of waste, water, electricity and other services. Although we see widespread private sector provision of basic services in a diffused system of local governance[2], it is unsurprising that the United Nations Sustainable Development Goals explicitly require local governments to ensure access for all to adequate, safe and affordable housing and basic services (SDG 11.1)[3]. In a similar vein, the United Nations New Urban Agenda envisages cities and local governments that provide universal access to safe and affordable drinking water and sanitation, as well as equal access to quality services such as infrastructure, energy, and transportation (13(a) NUA).[4]

South Africa has world-class legislation. After the adoption of the *Constitution of the Republic of South Africa*, 1996 (the Constitution), a remarkable process of law and policy reform followed. This is particularly true for the law on issues that were the subject of, or cause for, neglect for as long as apartheid stayed intact. The constitutional status of local government changed significantly after 1996. We reflect here on service delivery law reform over the past 30 odd years, focusing on the expectations that were cemented in law. This reflection is done in the context of the current widespread service delivery challenges and absent consequence management.[5] We take a step back to comment on the possible reasons for the legal framework struggling to deliver. We also use this opportunity to reminisce the original policy vision of 'developmental local government' and what was envisaged in the late 1990s for fair, sustainable and equal future access to services of a high quality.

The constitutional dream: Adequate services for all

The Bill of Rights in chapter 2 of the Constitution, read with the extensive constitutional provisions on cooperative government (chapter 3) and a local government dispensation characterised by autonomy and extensive authority, confirms a constitutional dream of sufficient basic and other services for all. In other words, the Constitution promises people rights in relation to adequate housing, access to water, and work and living environments not harmful to human health or well-being. These rights translate into duties of inter alia municipalities, as further informed by the objects of local government which include, for example, the provision of services to communities on a sustainable basis, promoting socio-economic development, protecting the environment, and providing democratic and accountable governance.

In a striking judgement some years ago (*Joseph and Others v City of Johannesburg and Others*)[6], the court went as far as using existing constitutional and legislative provisions that oblige local government to provide basic services to community residents to establish a "public law right" to receive electricity. The court held that although City Power had a contract with the landholder of Ennerdale Mansions, a flat of 44 blocks, it could not terminate electricity to the residents without complying with the procedural requirements of the right to administrative justice. The Court held that the termination of electricity supply to Ennerdale Mansions was unlawful and that residents were at least entitled to a pre-termination notice that would have given them an opportunity to make representations and find suitable solutions to the problem. The Court reasoned that the Constitution and legislation established a "special cluster of relationships" between municipalities and their residents, concretised by the bundle of public responsibilities that municipalities owe their residents in terms of the Constitution and legislation.

The spirit of the Constitution and the much larger transformative project it hopes to deliver suggest that sustained delivery of services has a big role to fulfil. The dream embodied in our Constitution is for people to live lives marked by human dignity, equality and respect for life. It is for this important reason that we grasp the consequences of failing service delivery. It has a direct bearing on human health, well-being, dignity, equality and may even put lives and ecosystems at risk.[7] In the bigger scheme of things, failing services also pose direct risks for the integrity and health of a still maturing democracy. Ordinary people lose faith in democratic processes and in representative leaders and agents of the state when the most basic of their needs remain unmet.[8] When service delivery protests erupt, one is often confronted with claims about unresponsiveness and lack of accountability among political leaders and authorities.[9] These sentiments fly in the face of the Constitution which explicitly commands that all constitutional obligations must be performed diligently (section 237).

It would be deceiving not to mention the flip side of this constitutional dream. The Bill of Rights repeatedly refers to 'everyone', which means that few people, regardless of their legal status, would be excluded from the entitlements in the Constitution. An exception is the political rights reserved for citizens, such as the right to vote or establish a political party. The inclusive constitutional protection understandably creates an inviting

environment – also for people beyond South Africa's borders. While we do not attempt to deliberate this point, it should be mentioned that services must be delivered to all and that in a country with exceedingly high rates of unemployment, urbanisation and upwards population growth trends, this understandably creates resource-related and other governance challenges.[10] By the same token, local government is not solely responsible for service delivery (or its failure). The democratic constitutional dispensation ushered in a government consisting of three spheres (national, provincial and local) and many line functionaries – all of which are expected to cooperate and collaborate (chapter 3 of the Constitution). This constitutional command to work together is now being put to the test with the Government of National Unity that came about after the 2024 national elections.

The policy vision: Local government for service delivery and more

A white paper setting out the policy direction for a specific issue runs the risk of becoming redundant and insignificant when an enforceable law on the topic comes about. We argue that an immensely meaningful vision for local government was captured in the 1998 White Paper on Local Government.[11] However, between the adoption of a suite of local government legislation and extensive judicial interpretation of local government powers and functions, much of the original conceptualisation of 'developmental local government' went missing – not only in the general narrative about service delivery and the duties of municipalities, but also, in the deeper understanding of what sustainable and equitable service delivery, firstly, should be and, secondly, may achieve in a country such as South Africa which aspires to be a healthy and thriving democracy.

The White Paper made it clear that, after apartheid, local government would have an extremely important role to fulfil. It declared local government as a key development agent tasked with much more than mere service delivery. Yet, service delivery was explained in the White Paper as to command dedication to the eradication of service delivery backlogs and the provision of services to all, in line with principles such as quality, accessibility, accountability, sustainability and value-for-money (White Paper 1998, Section F). One of the shortcomings of the White Paper is that it could not anticipate how poverty would deepen over time with a knock-on effect on the revenue raising of municipalities and the socio-economic conditions in which people would continue to live for decades longer. Some years later, the White Paper's overarching vision for service delivery would inform what became the statutory definition of basic municipal services, namely: "a service that is necessary to ensure an acceptable and reasonable quality of life, and if not provided, would endanger public health or safety or the environment" (section 1 of the *Local Government: Municipal Systems Act 32 of 2000*).

Many policies on development and local governance followed the White Paper, but perhaps the most insightful for present purposes is the Integrated Urban Development Framework of 2016 (IUDF).[12] This Framework reminds us that "despite significant service delivery and development gains since 1994, apartheid spatial patterns have largely not been reversed" and that, in a more forward-looking fashion, the provision of municipal services must now focus on "connected, resource-efficient public services, such as efficient energy, waste and water systems, street lighting, technology, and smart grids"[12].

The national policies on local government speak of an awareness of the importance and vast scope of its service delivery duty. This duty may, however, not be isolated from various other challenges, including the ongoing state of spatial disparity and injustice as well as the impacts of climate change and energy poverty, among other pressures. The persisting spatial inequalities across urban South Africa are a cause of great concern from environmental sustainability and service delivery accessibility perspectives.[13]

Statutory law turned on its head: Developments between 1996 and 2024

Local government operates in a highly regulated space and the impact of this for service delivery and good local governance has been lamented before.[14] South Africa has seen a flurry of local government legislation being rolled out from the year 1998 onwards, and today a suite of legislation applies to the structure, powers and daily functioning of municipalities. From a service delivery perspective, virtually all these laws find application, ranging from the *Local Government: Municipal Systems Act 32 of 2000* to the *Local Government: Municipal Finance Management Act 56 of 2003* and the *Local Government: Municipal Structures Act 117 of 1998*. These Acts are, however, only a drop in the ocean of statutes that direct the way in which municipal services should be provided in South Africa. Some of the other laws of direct relevance include the *National Environmental Management: Waste Act 59 of 2008*, the *Water Services Act 108 of 1997*, the *Spatial Planning and Land-Use Management Act 16 of 2013* and the *National Environmental Management: Air Quality Act 39 of 2004*. There are plenty more. Besides new laws, some apartheid-era legislation, which is often incompatible with the new constitutional scheme, still applies to municipalities in the building sector, for example.

Some of the legally relevant problem(s)

Despite (or perhaps because of) a strong law and policy framework for local government, legally relevant problems around service delivery persist. We briefly look into a few of these.

Although the legal framework for the newly proposed District Development Model (DDM) is emerging[15], the autonomy of municipalities[16] means that, although they are expected to participate in this developmental model, they cannot be forced to be part of it. In 2019, President Ramaphosa launched the DDM model as a potential panacea to local government's service delivery woes. The aim was to improve integration in planning and budgeting across the spheres of government through 44 district municipalities and eight metros. The DDM aspires to ensure that there is one plan and one budget that addresses the development priorities for each metro and district municipality in the country. The regulations of 14 May 2024 provide an intergovernmental and operational guide for the coordination of municipal intergovernmental development priorities in the context of the DDM, through established intergovernmental forums. The DDM model is still in its early years of implementation, and it is difficult to predict the eventual success of this model. Besides, given the strong constitutional autonomy of municipalities, the DDM model is voluntary, and municipalities cannot be coerced into it. It further does not affect the current legal distribution of powers between spheres of government or between district and local municipalities. Without a significant increase in fiscal allocation from national government, coupled with deployment of appropriately skilled persons, the DDM model may not make any significant difference, like many of the local government support programmes that came before it. An alternative approach could be for the Municipal Demarcation Board to conduct research to assess the viability of collapsing the 257 municipalities along the geographical boundaries of the current DDM model. This may radically reduce the number of municipalities that may use their autonomy to reject participation in the DDM and facilitate redistribution of existing expertise within new municipal boundaries. Potential concerns about the large size of municipalities along the DDM geographical boundaries are still applicable to the current DDM model.

Another major problem which has been regularly raised is the current model of local government finance. Despite their fiscal autonomy (sections 229 and 230A of the Constitution), most municipalities heavily rely on the national transfers, the Local Government Equitable Share (LGES) grant, which is guaranteed in terms of section 216 of the Constitution and given effect to discharge through the annual *Division of Revenue Act*. Weak revenue generation due to lack of viable tax bases and poor debt recovery hamper the financial viability of municipalities. There have been calls to government to revisit the LGES formula after its third and last revision in 2012/2013 to strengthen the fiscal viability of municipalities.[17] Given the scope of services that municipalities are expected to provide, and the decaying state of municipal infrastructure, national government needs to seriously consider this call for increased fiscal allocation to municipalities. Without this, many communities will remain in limbo.

Despite the legal framework to promote good financial governance in municipalities, significant amounts of money are lost through mismanagement and corruption.[18] Fiscal mismanagement as well as

limited skills and capacity in finance persist.[19] This situation has seen municipalities continue to return millions of unspent funds to National Treasury every year. In addition, corrupt procurement practices lead to wasted expenditure that could be used to deliver services. Attempts to recoup lost funds through judicial intervention have not been very successful for diverse reasons. Often, municipal officials approach courts to review and set aside corrupt procurement contracts when the implementation of projects has neared completion (*Buffalo City Metropolitan Municipality v ASLA Construction (Pty) Ltd*)[20], and the law is not clear on what percentage of the profits can be recovered from contractors that were part of the corruption. While contractors are generally allowed to keep profits earned from corrupt procurement deals once they have performed their tasks, courts have indicated a willingness to scrutinise excessive profits (see *Black Sash Trust v Minister of Social Development*[21]; *RAiN Chartered Accountants Incorporated v South African Social Security Agency*[22]; *Siyangena Technologies (Pty) Ltd v PRASA*[23]; *Govan Mbeki Municipality v New Integrated Credit Solutions (Pty) Ltd*[24].

Lastly, many municipalities across the country suffer severe shortages of critical (and scarce) skills.[25] The national government (e.g. the Department of Water and Sanitation) at times imports foreign engineers to service waste and water treatment plants.[26] This approach has been described as "bizarre" given that South Africa has qualified technical individuals and engineers who struggle to find employment.[26] Although the *Employment Equity Act 55 of 1998* (EEA) serves well-intentioned equity objectives (section 2 of EEA), there appears to be a need to relax rigid adherence to affirmative action targets set in terms of the Act where there are severe shortages of skills needed by municipalities. The position adopted by government has driven some with critical skills to open consultancies that milk municipalities by excessively charging for their services, often on basic matters that municipalities ought to easily handle.[19,2419] Instead of importing engineers and other expertise, the government should consider relaxing employment equity targets in municipalities where there is need for critical skills. Dogmatic adherence to affirmative action targets in critical skills sectors will continue to negatively impact service delivery and the lives of South Africans spread across formal and informal living environments.

Courts speaking out

The judiciary indirectly guards over the state of service delivery in South Africa. High Courts, the Supreme Court of Appeal and the Constitutional Court in South Africa are uniquely positioned to ensure constitutional compliance by municipalities (see chapter 8 of the Constitution). They are obliged to declare municipal law or conduct (action) invalid to the extent that it is inconsistent with the Constitution and have wide remedial powers to make any order that is just and equitable (section 172 of the Constitution). It is against the backdrop of this watchdog role that courts have been regularly approached by communities and community organisations disgruntled with the failure of municipalities to discharge their service delivery duties. In certain cases[20,24,27-31], courts have repeatedly expressed frustration with the failure of municipalities to manage and maintain critical municipal infrastructure, provide basic services to communities in a sustainable manner, stop the deliberate pollution of rivers with untreated sewage, meaningfully facilitate community involvement in local governance, prudently manage their fiscal affairs, the self-serving nature of some municipal officials, municipal disrespect of the rule of law and court orders, and the lack of sufficient coordination and cooperation between municipalities and other spheres of government. In response to these failures, and in line with their wide remedial constitutional powers, courts have used structural interdicts to supervise municipalities in order to ensure that they comply with their orders, ordered the imprisonment of municipal managers for failure to comply with court orders, ordered provincial and national government to intervene in failing municipalities, nullified parts of procurement contracts which enabled service providers to make excessive profit at the disadvantage of rate payers, and enabled a community-based organisation to take over and manage water works and sewerage infrastructure in places such as Koster and Swartruggens in the North West Province. Often, some court orders, including structural interdicts, do not lead to timely and/or full compliance with municipal duties. This points to the limits of judicial power to effect transformative change in the municipal service delivery domain.

Although this approach of the courts has often attracted criticism for violating the separation of powers doctrine, they have exercised remedial powers expressly conferred by the Constitution to ensure municipal legal compliance in instances of abysmal local government failures. Outside of courtrooms, the level of despondency has often forced community residents and taxpayer associations to illegally take over collapsing municipal infrastructure and provide municipal services such as road infrastructure to restore their towns.[27] These acts may be seen as manifestations of civil disobedience and cannot be condoned, whilst they display a desperate attempt among ordinary citizens to ensure *some* access to *some* basic services for *some* people in the short term.

Conclusion

South Africa has world-class legislation. So what then should we make of a situation in which the law struggles to deliver? What does it suggest for South Africa that our remarkable legal framework, embedded in the Constitution, but fails to deliver on the inherent and explicit promise of adequate service delivery? Various interdisciplinary and multidisciplinary research projects would be required to explore all aspects of this complicated question, as it has layers reaching into politics, economics, finance, public and private governance, as well as education, investment and the impacts of global environmental change.

At a minimum, it should be questioned whether we are faced with a legal design error spanning the period 1996 to 2024, or whether the law has been putting too much emphasis on state responsibility *vis-à-vis* that of other actors, such as industry and the financial sector. We should also ask whether some areas of the law – e.g. municipal finance management law, environmental law or law aimed at redress in the labour sector – stifle progress with service delivery because of stringent procurement and other statutory requirements and administrative procedures. The dynamic nature of 'the law' in these areas also poses a serious challenge to municipalities. Before municipalities adapt to a (new) law, changes are already anticipated, which constantly changes the goalpost. One example is the recent *Climate Change Act 22 of 2024* which requires of metropolitan and district municipalities to "undertake a climate change needs and response assessment" to be followed by a municipal "climate change response implementation plan" (chapter 3 of the Act). These instruments will no doubt have an impact on service delivery considering the exposure of municipal infrastructure to climate impacts.

By the same token, we should ask if non-compliance with the law is addressed with the necessary vigour given the desperate need for improved service delivery in a constitutional democracy with *three* autonomous, but interrelated, government spheres sharing the immense responsibility of cooperating towards and delivering long-term development and prosperity. The ability of service delivery law to deliver depends on the understanding that, if one sphere of government fails, the entire government fails. This rings more true than ever in the recently constituted Government of National Unity. Fair, sustainable and equal access to services of a high quality in a country composed of a population of now more than 63 million people, requires the utmost dedication of every single political office bearer, administrator and government official in each of the relevant departments spread across the national, provincial and local spheres of government.

Acknowledgements

The contribution of O.N.F. is based on research supported by the National Research Foundation (NRF) of South Africa (grant number 115581). The opinions, findings and conclusions or recommendations expressed in this publication are that of the author(s), and the NRF accepts no liability whatsoever in this regard.

Declarations

We have no competing interests to declare. We declare no AI or LLM use. Both authors read and approved the final manuscript.

References

1. Du Plessis A. A role for local government in global environmental governance and transnational environmental law from a subsidiarity perspective. Comp Int Law South Afr. 2015;48:281–316.
2. Pieterse M. Urbanizing human rights law: Cities, local governance and corporate power. German LJ. 2022;23:1212–1255. https://doi.org/10.1017/glj.2022.77
3. United Nations Sustainable Development Goals [homepage on the Internet]. c2015 [cited 2024 May 17]. Available from: https://sdgs.un.org/goals
4. UN-HABIT. United Nations New Urban Agenda. c2017 [cited 2024 May 17]. Available from: https://unhabitat.org/about-us/new-urban-agenda
5. Makamu T. The failure to spend the Municipal Infrastructure Grant and the lack of consequence management. Local Government Bulletin. 2024:19. Available from: https://dullahomarinstitute.org.za/multilevel-govt/local-government-bulletin/archives/volume-19-issue-2-june-2024/the-failure-to-spend-the-municipal-infrastructure-grant-and-the-lack-of-consequence-management
6. *Joseph and Others v City of Johannesburg and Others*. 2010. (3) BCLR 212 (CC) .
7. Chamberlain L, Khunou K, Masiangoako T, Potter A. Makana Local Municipality: Provincial intervention in a municipal crisis. Johannesburg; SERI; 2020. Available from: http://www.seri-sa.org/images/SERI_Makana_case_study_FINAL_WEB.pdf
8. Wright J, Dube F, Du Plessis AA. Judicial enforcement of mandatory provincial interventions in municipalities in South Africa. VRÜ Verfassung und Recht in Übersee. 2022;55(1):105–125. https://doi.org/10.5771/0506-7286-2022-1-105
9. Stoffels M, Du Plessis AA. Piloting a legal perspective on community protests and the pursuit of safe(r) cities in South Africa. South Afr Public Law. 2019;34(2):1–26.
10. Fuo O. Nativism in South African municipal indigent policies through a human rights lens. Law Democr Dev. 2020;24(1):271–317. http://dx.doi.org/10.17159/2077-4907/2020/ldd.v24.12
11. RSA. White Paper on Local Government [document on the Internet]. c1998 [cited 2024 May 17]. Available from: https://www.cogta.gov.za/cgta_2016/wp-content/uploads/2016/06/whitepaper_on_Local-Gov_1998.pdf
12. RSA. Integrated Urban Development Framework [document on the Internet]. c2016 [cited 2024 May 17]. Available from: https://iudf.co.za/pdf_downloads/2016-integrated-urban-development-framework/
13. Van Wyk J. The law on planning and the environment. In: King ND, editor. Fuggle and Rabie's environmental management in South Africa. 3rd ed. Cape Town: Juta; 2018. p. 1135–1176.
14. Steytler N. The strangulation of local government. J South Afr Law. 2008; 3:518–535.
15. Regulations Framing Institutionalisation of District Development Model in Terms of Section 47(1)(b) of Intergovernmental Relations Framework Act, 2005. Published in Notice 2486 in Government Gazette No 50645 of 10 May 2024. Available from: https://www.gov.za/sites/default/files/gcis_document/202405/50645gen2486.pdf
16. Fuo O, Pallangyo DM. A comparative legal analysis of local government's autonomy in Tanzania and South Africa. J Comp Law Afr. 2023;10(2):1–39.
17. SALGA, National Treasury, Department of Cooperative Governance. Local government equitable share Formula Review. Discussion Paper 1: Proposed principles and objectives of the Local Government Equitable Share Formula. c2012 [cited 2024 May 17]. Available from: https://www.salga.org.za/Documents/Municipalities/Municipla%20Focus/Municipal%20Finance/B_-LGES-Discussion-Paper-1---Principles-and-Objectives.pdf
18. Fuo O. Funding and good financial governance as imperatives for cities' pursuit of SDG 11. In: Aust H, Du Plessis AA, editors. The globalisation of urban governance. London: Routledge; 2019. p. 87–107.
19. AGSA. Consolidated General Report on Local Government Audit Outcomes MFMA 2021-22 [document on the Internet]. c2023 [cited 2024 May 17]. Available from: https://www.agsa.co.za/Portals/0/Reports/MFMA/2021-22/MFMA%20Report%202021-22%20FINAL%20-%2031%20May%202023.pdf?ver=2023-05-31-072950-280
20. *Buffalo City Metropolitan Municipality v ASLA Construction* (Pty) Ltd. 2019 (6) BCLR 661 (CC).
21. *Black Sash Trust v Minister of Social Development (Freedom Under Law intervening).* 2017 (5) BCLR 543 (CC).
22. *RAiN Chartered Accountants Incorporated v South African Social Security Agency.* 2021 (11) BCLR 1225 (CC)
23. *Siyangena Technologies (Pty) Ltd v PRASA and Others* (487/2021) [2022] ZASCA 149 (1 November 2022).
24. *Govan Mbeki Municipality v New Integrated Credit Solutions (Pty) Ltd* (unreported) case number 121/2020; of 7 April 2021
25. Local Government Sector and Training Authority. Sector Skills Plan 2024/25. 2023 [cited 2024 May 17]. Available from: https://lgseta.org.za/wp-content/uploads/2024/06/SSP-24_25.pdf
26. Papenfus G. Failing municipalities, Cuban engineers and the ANC's race policies. c2021 [cited 2024 May 17]. Available from: https://library.neasa.co.za/failing-municipalities-cuban-engineers-and-the-ancs-race-policies/
27. *Kgetlengrivier Concerned Citizens and Another v Kgetlengrivier Local Municipality and Others* (UM271/2020) [2020] ZANWHC 95 (18 December 2020).
28. *Mafube Business Forum and Another v Mafube Local Municipality and Others* (1969/2021). [2022] ZAFSHC 86 (28 April 2022).
29. *Unemployed Peoples Movement v Eastern Cape Premier and Others.* 2020(3) SA 562 (ECG).
30. *Featherbrooke Homeowners Association NPC v Mogale City Local Municipality & Others.* (GSJ) (unreported) case number 11292/20 of 25 January 2021.
31. *Minister of Water and Sanitation v Msukaligwa Local Municipality and Others* (unreported) case number 4860/2022 of 20 September 2023.
32. *Kamiesberg Local Municipality and Another v Koingnaas Belastingbetalers-vereniging and Another* (2609/2021) [2021] ZANCHC 69 (31 December 2021).

AUTHORS:
Gordon Inggs[1]
Delyno du Toit[1]

AFFILIATION:
[1]Data Science Branch, City of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Gordon Inggs

EMAIL:
gordon.e.inggs@ieee.org

HOW TO CITE:
Inggs G, du Toit D. Data analysis is part of a better feedback loop for local government. S Afr J Sci. 2024;120(11/12), Art. #19198. https://doi.org/10.17159/sajs.2024/19198

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
data analytics, local government, service delivery, systems analysis, feedback loops

PUBLISHED:
30 August 2024

Data analysis is part of a better feedback loop for local government

Significance:

This Commentary addresses how local government in South Africa should organise its decision-making processes to best serve their residents. We illustrate and discuss how system analysis can be applied to local government decision-making in the South African context, in the vein of calls for "evidence-based decision making". It also proposes how data analytic work may be incorporated into the local government system, both conceptually, as well as in practical terms, through organisational and cultural interventions.

Throughout South African society and beyond, it is widely accepted that local government is both critical to the delivery of basic services, and the site of some of the most significant failures thereof. There are many astute analyses of this in both the popular and academic press, and we do not feel that we have much to add to these diagnoses. What we believe we can offer is something of a remedy. Our suggestion does not necessarily directly address the reasons for service delivery failure, but rather, in our experience, provides the platforms or fora for doing so.

We argue that a relatively minor change in how we organise the work of local government, namely the introduction of analytic feedback processes, could have a transformative effect. We provide background on the relevant systems theory before expanding on two of the modes of feedback that we have observed in local government. Finally, we make some practical suggestions on how this could be realised.

Feedback loops

Figure 1 is a representation of one of the most important ideas in what was once called cybernetics theory, and is more commonly today called systems analysis: the humble feedback loop.

What we mean by *implementation* in this diagram is intentionally very broad. It is the occurrence of a defined process or behaviour of a system; for example, the balancing of power across the four engines of a quadcopter drone, or the monthly billing of residents of a city for municipal services.

This might seem very abstract, but a surprising number of quite varied activities can be mapped into this model, for example:

- A sports team over the course of a season or tournament, playing multiple matches.
- A recommendation algorithm for an online shop that decides what products to present to a user who has logged in.
- A person learning a musical instrument for the first time, taking periodic lessons with an experienced teacher.
- A car journey in which a driver is making decisions on a continuous basis, managing the functioning of the vehicle as well as guiding it to a destination.

For the purposes of this discussion, we define the components of the loop as follows, using the example of the car given above:

- *input* – anything that is required for the system in which we are interested to function. This might be a practical need, e.g. fuel needed for the engine of the car, or an informational need, e.g. the signal from the key being turned in the ignition.
- *decision* – the point at which some sort of intelligence, human or machine, is applied to the inputs, to determine the behaviour of the system. For our car, this would be the decisions made by the driver of the car, in response to the various signals received from the driver's dashboard, as well as observations of the outside world.
- *implementation* – the occurrence of a process or behaviour of a system of interest. In our example, this would be the actual running of the engine, and in turn, its ability to cause the car to move.
- *effects* – what results from the occurrence of the process or system in which we are interested, i.e. the movement of the car resulting from the engine functioning. Other effects would include the emission of exhaust fumes.
- *feedback* – the capturing or measurement of the implementation. Of course, nothing can be measured perfectly, so this is a view on what occurred during, and as a result of, the implementation. In our car, this would be the information conveyed by the dials on the dashboard.

Any process that has some sort of cycle, or is repeated, and where those participating have memory of previous iterations, can be thought of as a feedback loop. And, although intuitive to understand, the introduction of feedback from the outputs of the system, to the decision point, results in dynamic behaviour that makes predicting future outputs of the system surprisingly difficult. We call such systems *complex* and this sometimes surprising behaviour *emergent*, as it arises from the interaction of various elements of the system, as well as past behaviours of the system itself, as opposed to the particular attributes of those elements in isolation.

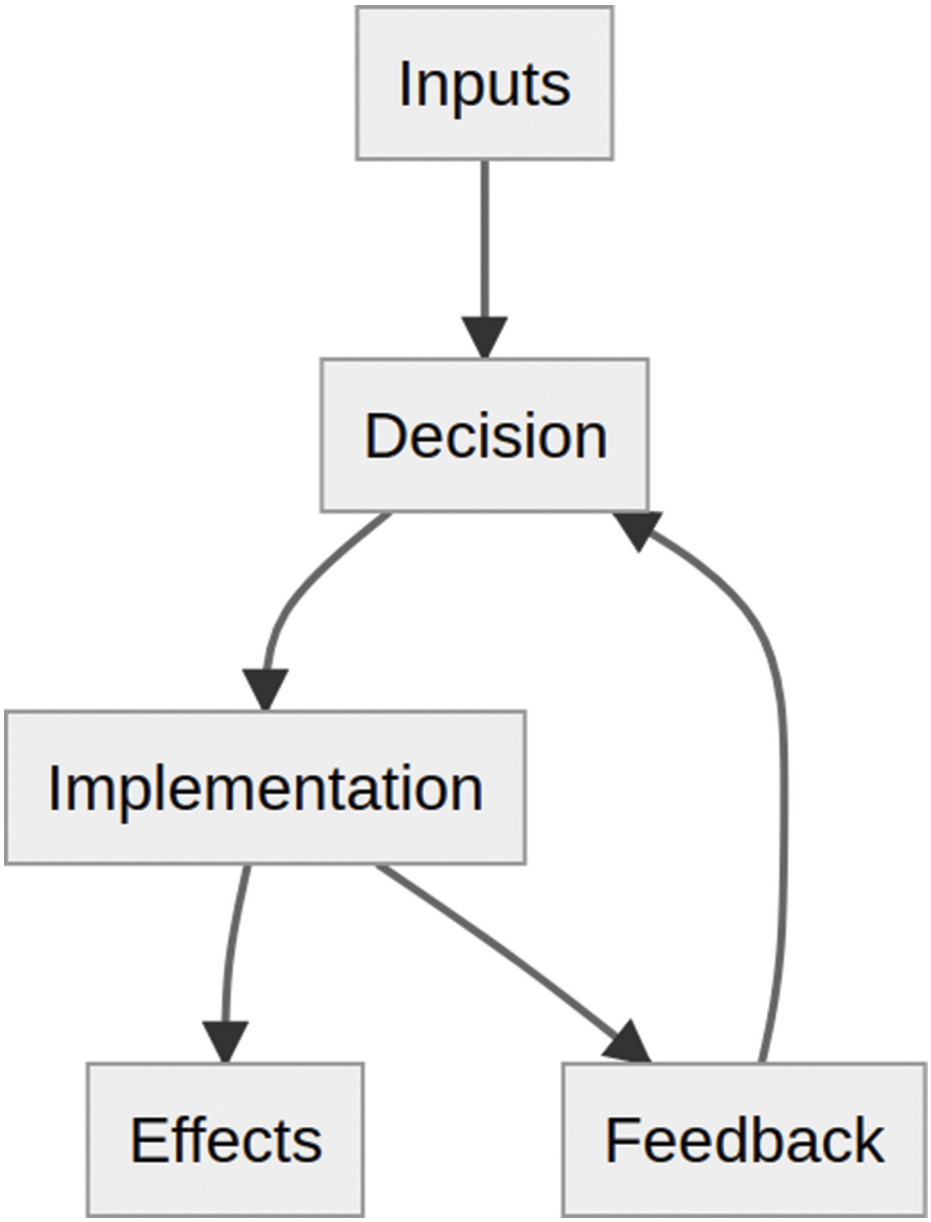


Figure 1: Simple system overview with feedback loop.

Generally speaking, if one is seeking stability and consistency in the outputs, we think about a *negative* feedback loop. This means that the feedback moderates or adjusts the inputs such that the outputs are kept at a constant level. A *positive* feedback loop amplifies inputs, such that they grow until they reach some sort of limit, such as exhausting the resources available.

Relationship to digital data technologies

Before moving onto how this modelling approach relates to local government, we discuss the relationship with the dramatic growth in digital computing capabilities over the last 100 years, the corresponding growth in data available for processing (often referred to as 'big data'), as well as techniques such as statistical computing and machine learning (and its related subfield artificial intelligence) for gaining insight from this data.

It is no coincidence that systems analysis emerged and has developed in parallel with these socio-technical trends, and indeed, many theorists such as Jay Forrester and John Von Neumann have played an important role in developing both. However, it is important to understand systems analysis as a distinct set of modelling techniques for reasoning about systems, whereas these digital technologies are often the means by which these systems are realised.

The availability of digital technologies, and more importantly the capability to implement and operate them within an organisation, often greatly increases the visibility of the systems in question, and, in this way, makes obtaining and putting the insights of systems analysis into action easier. But it must be stressed that these technologies ease the use of system analysis as opposed to being a prerequisite. No technology beyond a pen and paper is required to model a system, and nothing beyond a meeting with the appropriate people is required to put these insights into action.

There is sometimes a worrying tendency in government to conflate the use of these technologies (often in the form of 'digitalisation' programmes) with making better decisions, or progress more broadly. Our experience is that this is a very dangerous and potentially expensive mistake to make – having the right tools is no guarantee that a job will be done correctly.

Government decision-making as feedback loops

Implicitly, feedback loops are everywhere, and indeed are inherently part of how we all make decisions, both conscious and unconscious. After all, we often repeat activities, and have memories of those past attempts. However, much of this is implicit – only the very disciplined keep a detailed

record of past attempts, and repeat the learning process. Often, without thinking about it, we incorporate our experiences and beliefs about how the world should be, when deciding what we are going to do next.

However, in government, there are many reasons why we keep more track of the decisions we make, and the information considered when making those decisions. Often those decisions relate to the allocation of scarce resources and so efficacy and efficiency are desirable, or those decisions relate to the interpretation and application of legislated norms, and so would be subject to later scrutiny for accountability reasons.

So, we suggest that government itself, and particularly local government, being a socio-technical system responsible for the delivery of many basic services (at least, in the South African context), can be modelled usefully as systems with many feedback loops. There are often clear decision points, from which a variety of actions result, which in turn, result in various effects for those governed. There are also established ideas of measuring these effects, and this measurement is often seen as part of the work of government.

The alternative which we have encountered is that, when trying to describe these systems and their behaviour, language and concepts are simply borrowed from private sector organisations such as large corporations. While there are a lot of similarities, in that these are organisations of people with a common identity and nominally shared goals, public sector organisations have considerably more, diffuse objectives, in contrast to the private sector's single one. We believe that by using the language of systems modelling, and in particular of feedback loops, the often surprising, emergent behaviour of local government organisations can be better explained.

It is from attempting to use the measurement of the effects of past decisions, as well as the inputs at the time, to help make better future decisions, that we arrive at the idea of *evidence-based decision-making*. Or put more directly, if we hope to achieve what we plan, we have to learn from the past.

Feedback modes

There are indeed many feedback loops operating at all levels of local government, particularly in relation to service delivery, such that trying to map out and understand all of these would be a considerable undertaking. However, the mechanics of these loops are not all equal. As the final part of our discussion, we would like to examine a particularly pathological feedback typology that we believe inhibits service delivery, and we propose an alternative that we believe enhances it.

To examine these two typologies, we ignore the inputs and effects for now, and just focus on the decision-implementation-feedback components, as given in Figure 2.

The reporting feedback process

Figure 3 is a depiction of a very simplistic version of the feedback process, and is indeed one that we have often observed in local government.

In this process, data or information are gathered, and then tabulated into a pre-ordained structure. That populated rubric, matrix or *report* is then submitted to a decision-maker, on the understanding that the information is required for their decision-making. This is almost entirely a *feed-forward* process, with little or no information flowing back to those providing the information by design (it might happen informally, or outside this system).

However, often in this type of feedback loop, which is large, spanning beyond the local government organisation, the *report* is submitted to an oversight body for scrutiny. The mandated frequency loop is also often quite slow, spanning multiple months or even years from information being collected to being submitted for scrutiny.

This version of the feedback loop has its merits:

- It is simple to implement.
- It does not require special skills or knowledge to operate.

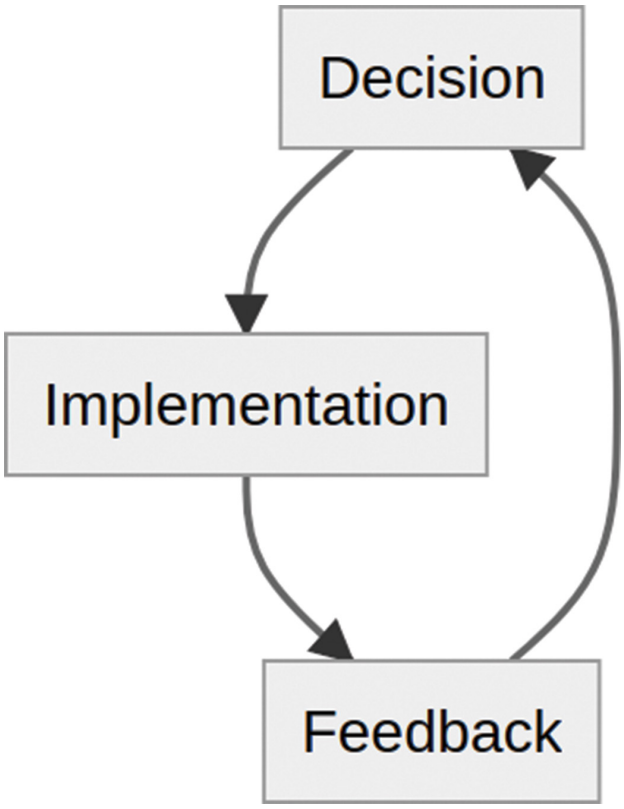


Figure 2: Overview of feedback loop components.

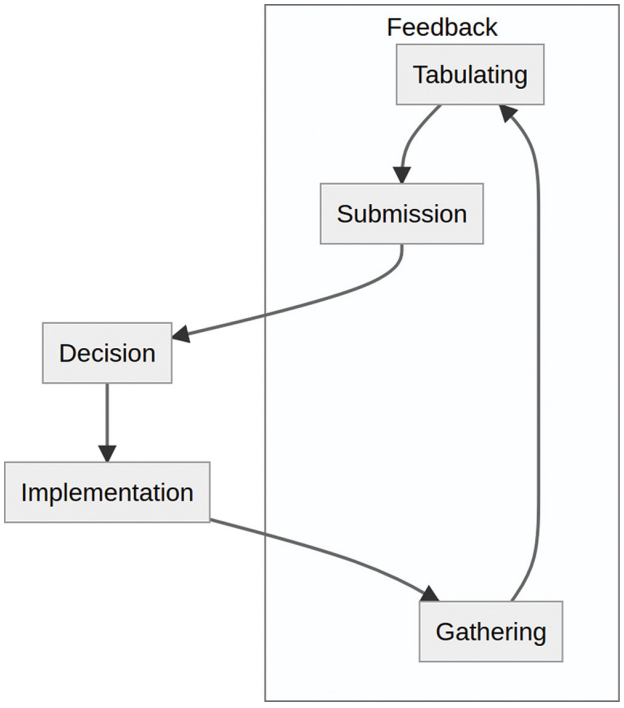


Figure 3: Feedback loop in the simplistic, yet widely observed 'reporting' mode.

As a result, it is often administrative staff who provide this information on a routine basis, usually by extracting it from transactional systems. These data are then typically manually manipulated into the form required, before submission. Other than ensuring the reports produced are in accordance with the predetermined structure, there is little to no expectation that these staff will overly concern themselves with the content of what is being produced.

In our experience, this feedback model has some serious problems:

- It is inflexible – an unfamiliar or unanticipated occurrence will fall outside the scope of what is being measured, or cannot be coerced into the predetermined structure. Dramatic events and the accompanying upheavals, such as the Cape Town drought of 2017, COVID-19 pandemic and July 2021 riots, are all examples of situations in which we observed that conventional reporting mechanisms became defunct, either returning readings that made no sense, or could not be completed under the circumstances.
- It is brittle – the nature of the system under observation only has to change slightly, and often the monitoring or evaluation framework has to be completely reworked to handle the new structure. An example is how the allocation of free basic services has changed over the last few years, and each change has upended reporting on the provision of these services.
- The focus is on process, not content – because those involved in creating the information are alienated and disconnected from the accompanying decision-making, their focus, quite understandably, is on getting this work out of the way as quickly as possible. At times, we have seen examples of reporting at a very high level of the consumption of services that are physically impossible, beyond the physical capacity of infrastructure to provide it.

As an aside, it is also telling that digital technology is mostly incidental to this feedback mode. Indeed, as far as we can tell, all that computers offer this reporting feedback model is a slightly more convenient way to produce and format tables of numbers.

Yet, by far the worst problem with this model is that there are bad incentives to participate in good faith. Because the content of the reports provided by the feedback mechanism is in turn often used to evaluate those running the feedback process itself, there is considerable pressure to submit the 'right' number, or at least the set of results which are less likely to warrant further scrutiny.

And even with processes in place to ensure the integrity of the information being submitted, admirably undertaken by entities such as the Auditor General, there is considerable risk in highlighting any limitations or failures of the system. This is because those most familiar with the system under observation are those gathering the relevant information, and are also those operating it, so are likely to be held accountable for its failures.

The analytic feedback process

The refinement that we propose is depicted in Figure 4, and simply incorporates a data analytic function into the feedback process.

In this model, data are gathered and analysed by someone or a group of people with data analytic training and skills. There is then a process to contextualise this analysis, which we think is best done by talking to relevant operational staff as well as those affected by the system (e.g. residents), and observing the processes described by the data in the real world. Finally, this contextualised analysis is reflected upon, usually in some sort of group setting such as an operational review meeting, with the relevant decision-makers present. This feedback process is a closed loop, with reflection on previous contextualised analysis driving further analysis.

This process addresses two of the main concerns of the 'reporting' process:

- It is flexible – as and when there are changes made to the system under observation, through the analytic and contextualisation activities, the feedback process can be adjusted accordingly.
- It provides better incentives to those involved – by defining the closed analytic cycle, it reframes the roles of those involved as seeking improved knowledge of the system and its behaviours. Hence, we believe this translates into different behaviour when strange or abnormal behaviours are detected. These anomalies are then opportunities for demonstrating the skill of those involved, and growing their knowledge of the domain. Knowledge of the system and domain thus becomes the mechanism by which they derive status.

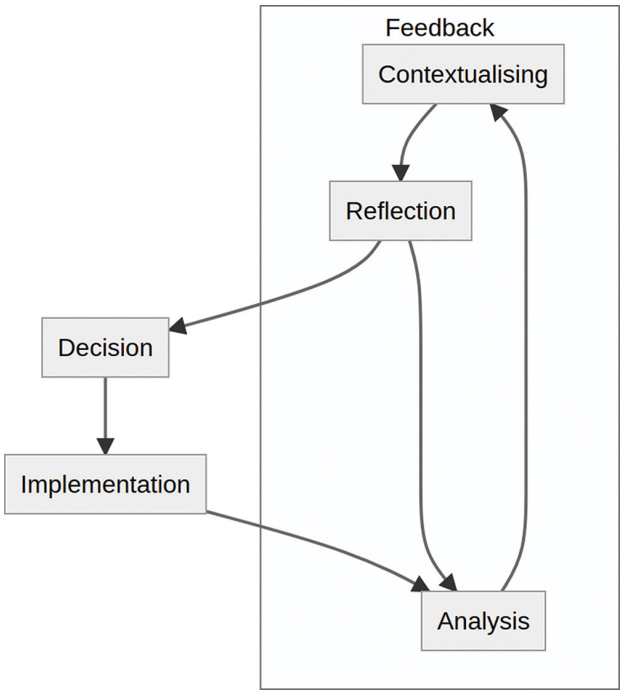


Figure 4: Feedback loop in our proposed 'analytics' mode.

We have seen the potential of such an approach in the service delivery turnaround programme run at the City of Cape Town in 2020 and 2021. Initiated to address the effect on service delivery caused by the COVID-19 pandemic, this organisation-wide effort used a data analytic-driven feedback loop to drive considerable improvements in service delivery.

Beyond the improvement seen in key performance metrics, such as the reduction in the backlog of service delivery related requests, as well as an improvement in the resolution time of requests, a marked feature of the programme was how often it transcended traditional organisational silos or structures. Beyond operational decision-making, functional areas such as human resources, vehicle and materials provisions, and contract management were all considered in driving overall system improvement.

Self-organisation, which is the ability of components of a system to improve coherency among themselves and become more efficient without intervention, has long been recognised as a key feature of robust systems. We believe that one of the great advantages of this data analytic driven feedback loop is that it encourages self-organisation in the delivery of basic services.

However, this data analytic model does require two significant commitments:

- Skilled people to participate – local governments, particularly those in rural areas, will struggle to attract and retain people with such a skillset.
- A degree of psychological safety, in which anyone involved can talk openly about the problems they observe, and pursue them without concern of reprisal. Traditional hierarchies and silent deference to age need to be suspended for these feedback loops to function.

In our experience, achieving both of these, particularly given the constraints of the South African local government context, is extraordinarily challenging, but highly rewarding.

Conclusion

We have described feedback loops, and how they prove a useful tool to think about decisions made within government in the South African context. In particular, in their potential to explain the often counterintuitive behaviour of government organisations, they show their value. We also described two models of feedback loops appropriate to supporting service delivery, one widespread and ineffective, and the other costly, but highly potent.

In closing, it is also worth reflecting that even thinking of government organisations as systems, or a set of tools for delivering services, is itself a somewhat radical notion. Instead, we too often view our bureaucracy as an immutable structure that cannot tolerate alteration or adjustment. Regardless of whether you accept our framing and suggested remedy, you would surely agree that we need to be willing to try to do better.

Recommendations

These are practical suggestions that we believe would help towards creating and supporting *analytic* feedback loops:

- Map out the high-level feedback loops for the provision of each basic service provided by a local government organisation. Particularly important is understanding the latencies or time delays in the flows of information, as this determines the speed of the system's response.
- Improve the data analytic capability of the organisation. Our experience is that this skill has to be hired, as it is difficult to retrain existing staff, particularly those who are very comfortable with the existing reporting process. Analysts from third parties or service providers can possibly be used in the short term, but it is difficult for them to acquire enough context in the organisation to be truly useful.
- Intervene in the organisational culture to educate around and promote *curiosity* and *humility* as core organisational values. This should be done in pursuit of creating the relevant psychological safety required for good analytic work to occur.

Acknowledgements

A working version of this Commentary was presented to our colleagues in the Data Science branch, and many of their helpful questions and comments resulted in considerable improvements. More broadly, our colleagues across the City of Cape Town, particularly those with whom we collaborate, continuously pursue service delivery excellence in making Cape Town a City of Hope.

Declarations

We assert that this Commentary is the result of our own efforts. Furthermore, we declare that generative AI was not used in writing the body of this Commentary.

Meta's Llama 3.1-8 Billion Parameter Instruction tuned model was used to obtain suggestions for the statement of Significance. The following prompt was used, in addition to the body of the Commentary: "Please suggest a 100 word or less significance statement for this commentary, articulating why it is important beyond academia."

The returned content:

> *This commentary highlights the limitations of current feedback models in government service delivery and proposes a new analytic feedback process that can improve decision-making. Its significance extends beyond academia to inform policy reforms and organizational changes that can enhance public service effectiveness and accountability. By bridging the gap between data analysis and operational realities, this approach has practical implications for local governments seeking to optimize resource allocation and improve citizen outcomes.*

AUTHORS:
Maréve Biljohn[1]
Grey Magaiza[2]

AFFILIATIONS:
[1]Department of Public Administration and Management, University of the Free State, Bloemfontein, South Africa
[2]Center for Gender and Africa Studies, University of the Free State, QwaQwa, South Africa

CORRESPONDENCE TO:
Grey Magaiza

EMAIL:
magaizaG@ufs.ac.za

HOW TO CITE:
Biljohn M, Magaiza G. Reimagining local government service delivery through the quadruple helix. S Afr J Sci. 2024;120(11/12), Art. #18711. https://doi.org/10.17159/sajs.2024/18711

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
service delivery, quadruple helix, government, municipality, partnerships

PUBLISHED:
30 August 2024

Reimagining local government service delivery through the quadruple helix

Significance:

An effective service delivery praxis is an embedded requirement of public policy frameworks. The current service delivery failures, particularly at the local government level, have exposed public policy shortcomings and disrupted livelihoods in South Africa. At the core of an evidence-based policy-enabled service delivery regimen is a 'better life for all'. At a local government level, the integrated development plan of municipalities directs this 'better life' through service delivery priorities. By interrogating the service delivery and quadruple helix nexus, we provide future pathways to reimagine the country's local government service delivery praxis to achieve transformative aspirations and development futures.

Framing the quadruple helix

Regarded as an extension of the triple helix, which has widely been studied as an effective partnership model between academia, government and industry, the quadruple helix emphasises a fourth stakeholder to enhance innovative end-user service delivery experiences and collaboration.[1] The quadruple helix adds civic society to the three helices mentioned above to augment relational networks and advocacy for service delivery implementation. In the context of this Commentary, the quadruple helix encompasses the civic sector, industry, the public sector, and academia. In our local democracy, these stakeholders are central to keeping local government accountable regarding their legislative mandate. More recently, the accountability role of these stakeholders has expanded to co-producing services, finding alternative solutions to service delivery challenges, and taking responsibility for the self-delivery of public services. A quadruple helix creates a service delivery ecology embedded in effectively disseminating capacity, skills, expertise, and resources to address local government service delivery challenges to enhance service delivery. A quadruple helix is, therefore, suggested as an inclusive, dynamic and intertwined service delivery approach at the local government level.

Policy and the service delivery context

South Africa has been touted as a developmental state, meaning its resources support a state-led and inclusive development aspiration. The National Development Plan (NDP), which is the main policy vision of the state, evinces the need for a capable and developmental state to review and correct injustices of the past. The implementation of development policy in South Africa has remained consistent with an inclusive and justiciable policy pathway that has emphasised state-centric development plans. Beginning with the Reconstruction and Development, Growth Employment and Redistribution Strategy, Accelerated Shared Growth Initiative for South Africa, New Growth Plan and NDP, these plans have emphasised service delivery and macro-economic fundamentalism to drive societal transformation and livelihood transitions. The reality, however, has been that the triple challenges of poverty, unemployment and inequality have steadily increased since 1994.[2] Therefore, in the 30 years of democracy, this policy context has seemingly not yielded the desired impact on service delivery. This is attributed to various factors emanating from the institutional internal environment supporting service delivery and its policy context. These factors include a lack of good governance, sound financial management, instability in key administrative positions, state capture and policy compliance. Ordinary citizens feel this lack of policy directly impacts their everyday lives regarding health, social, economic, environmental, educational, and basic services.[3]

At a local government level, this 30-year democratic narrative does not look much different for South African municipalities. To this end, the local government landscape is characterised by policy failures, poor audit outcomes, corruption, a lack of financial accountability, ailing service delivery infrastructure, cadre deployment, and key officials who lack appropriate financial skills and competencies.[4] It is an unfortunate continuity in failure as these municipalities are instrumental in addressing the country's triple challenge through developmental local government mandates. Although significant progress has been made towards improving the triple challenge and its impact on communities, municipalities cannot stay ahead in addressing development challenges. Twenty-six years after introducing the 1998 White Paper on Local Government, which set the scene for developmental local government, it seems municipalities are moving further away from living this vision. Several factors have contributed to the policy and developmental distance. The most noteworthy is the institutional distress, which exacerbates this distance and emanates from challenges in its internal environment. Such challenges include staff turbulence, poor infrastructure maintenance and planning, corruption, political abuse of power, the inability to execute basic functions and service delivery protests.[4] A call for alternative development roadmaps rooted in a service delivery praxis that delivers a 'better life for all' through a quadruple helix nexus is suggested to narrow this developmental local government gap and reposition municipalities with their mandate.

Dysfunctional or stable – South Africa's state of local government

Before re-imagining South African local government service delivery, taking stock of the state of municipalities is important. On 15 November 2023, the National Ministry of Cooperative Governance and Traditional Affairs presented the State of Local Government Report to a Parliamentary Committee. The findings of this report indicate that out of the 257 municipalities (8 metropolitan, 44 district, and 205 local municipalities), 30 are stable, 54 are categorised as low risk, 107 as medium risk, and 66 as dysfunctional).[5] The dysfunctionality in these municipalities

originates from systemic factors such as (1) a weak administration, (2) poor maintenance of infrastructure, (3) poor management and collection of revenue, (4) inability to "perform on conditional grants", and (5) instability of councils.[5] The 66 dysfunctional municipalities represent a quarter of South Africa's municipalities. If their stability is not restored, they could become part of the 32 municipalities that have been placed under administration since 2016.

Placing municipalities under administration in the South African local government sphere is a phenomenon that has been around for a while. This practice supports restoring stability in affected municipalities. There are, however, some municipalities in which the administration has been instituted repeatedly and which relapse to their dysfunctional state after such support. The administration in municipalities occurs through invoking section 139 of the Constitution, which permits the provincial executive to intervene in municipalities.[6] As a last resort, such intervention could stem from the inability of a municipality to fulfil its legislative obligation, including the approval of a budget and revenue-raising measures that will give effect to a budget.[6] Secondly, a crisis in the financial affairs of a municipality impacts its ability to meet its financial obligations; it has led to a material breach to provide essential services.[6] The 32 municipalities placed under administration are spread across eight of South Africa's nine provinces, with the majority falling within the category of local municipalities and six district municipalities.[5] Of concern is that, in 2022 alone, administration was instituted in 13 municipalities, which is significantly higher than in previous years.[5]

The deteriorating state of local government service delivery in some South African municipalities, amongst which are the dysfunctional ones and those placed under administration, is merely symptomatic of dysfunctional institutional systems that underpin its service delivery governance. If the crux of this symptomatic manifestation, which is seen daily in a lack of or poor service delivery, is not addressed, there will be a complete failure in these municipalities. These institutional systems are embedded in the local government service delivery system (social and technical), labour relations, financial management, and governance system of municipalities. Noteworthy regarding the functionality of these institutional systems is their recurring identification as challenges. In fact, a decade ago, this was identified in a previous State of Local Government in South Africa document by the Department of Cooperative Governance and Traditional Affairs.[7] To this end, the recurring nature of institutional challenges related to financial management, governance, service delivery, and labour relations has resulted in it being prioritised by two South African local government reforms: the Local Government Turnaround Strategy[8] (introduced in 2009) and the Back-to-Basics campaign[9] (introduced in 2013).

My business, your business, everybody's business

Every South African should ask: What does the state (dysfunctional or stable) of local government have to do with me? The stability of the 257 municipalities affects South Africans from all sectors of society. Regardless of the quality of services, it impacts people's livelihoods, businesses' sustainability, and municipalities' ability to attract and retain investment. The quality, sustainability and responsiveness of local government service delivery are the gateway to development in the towns and cities governed by municipalities. Moreover, it determines the state of development in municipalities. The worsening service delivery climate has seen an increase in costs for the delivery of some services as the private sector begins to occupy the space that is supposed to be led by the state; examples include the provision of health care, education, and security. Although some South Africans might be in an economic position to explore alternative service delivery arrangements, such as going off-grid, alternative water resources, and refuse collection, a significant number of our 60 million population are not in an economic and financial position to resort to alternative servicing. These citizens' lives will be most adversely affected by failing municipal and institutional systems that cannot deliver services sustainably.

Given the state of local government service delivery in the country, it is apparent that municipalities cannot save themselves. In addition to the intergovernmental relations system, the broader capacity of society has become more and more requisite to the future sustainability, responsiveness and resilience of local government's institutional systems that drive service delivery. Such capacity is located in the quadruple helix of society beyond the institutional framework of local government. Compared to the traditional role of service users, which are merely service recipients, this quadruple helix of stakeholders fulfils an unconventional but contemporary role where they are more central to decision-making processes regarding service delivery, resourcing and impact. In this reimagined service delivery context, the quadruple helix is involved in finding and implementing innovative solutions to the systemic institutional challenges, such as the local government service delivery system, labour relations, and financial management, that result in the current state of service delivery. These stakeholders support the local government service delivery system through their own and societal resources. Aligned with more contemporary service delivery approaches, the quadruple helix also participates in co-planning, co-designing, co-delivering and co-evaluating services.

Recent developments in the local government service delivery landscape show that citizens are becoming more organised in taking responsibility for some services that municipalities are supposed to deliver. Industry and higher education institutions likewise support collaborations towards repurposing municipal capacity through training and development initiatives. This demonstrates a willingness from society to ensure that local government's service delivery system remains responsive and redefines society's current and future role in South Africa's local government service delivery landscape. However, the extent to which regional governments are open and ready for this role – which requires navigating institutional cultures, bureaucratic systems and the dynamics of the political-administrative interface – is uncertain.

Reimagining capacity for service delivery in South Africa

We emphasise that service delivery is the cornerstone of transformation and should be considered a national priority. Such transformation starts at a local government level, where the direct impact of service delivery is felt. An effective service delivery approach must go beyond a mechanistic emphasis and highlight holistic and integrated foci that clearly articulate acceptable indicators to monitor transformation milestones. These transformation milestones should address the institutional and systemic challenges affecting service delivery systems. Several pathways to make this possible are suggested. The first pathway is to return to basics, entrenched in delivering quality basic services. The service delivery system profoundly impacts local economic development in municipalities and contributes to the gross domestic product of provinces.

Moreover, sustainable service delivery affects the ability of established and small businesses to survive in the current economic climate. A return to the fundamentals of basic service delivery is critical to ensure responsive, effective, efficient and sustainable service delivery. Integral to this is revisiting whether municipalities can deliver the complete basket of services aligned with their mandate. This might require reconfiguring its service delivery mandate. Second, is the geographical size of areas municipalities must serve within their resource constraints. Municipalities serving smaller geographical areas might be part of its future construction to ensure effective and efficient service delivery.

The second pathway should be embedded in policies that enhance the institutional systems of local government pertaining to its service delivery system, labour relations, financial management, and governance system. Concerning its service delivery system, policies should equally prioritise the technical and social sub-systems for equitable service delivery. Whilst the technical sub-system is concerned with policies, procedures and processes that determine service delivery, the social sub-system relies on the participation of the quadruple helix in the co-production of services.[10] One of the major areas for improvement of the current service delivery system is the poor integration of inputs from its social sub-system in the co-production (co-planning, co-design, co-delivery and co-evaluation) of services. The quadruple helix stakeholders could

significantly address some of the local government's pressing service delivery challenges through their skills, expertise, and resources.

Recognising that not only established or formal sectors of society should play a role in addressing challenges is essential. Entrepreneurs and small businesses are equally vital in identifying solutions to service delivery improvements through new business ideas. Some of these businesses are part of the diverse and large base of active and engaged entrepreneurs in the critical service delivery sectors. These entrepreneurs' and small businesses' roles in service delivery should be facilitated through the right policies at the local government level. Given the local government's skills and capacity deficit to deliver some service priorities, this small business and quadruple stakeholders could become critical partners in delivering services without municipal capacity. Where these and other businesses become a delivery mechanism for municipal services, their appointment should be based on technical competence and capacity to deliver the procured services. Creating an effective entrepreneurial and business support ecology is a critical enabler for a functional service delivery environment designed to fulfil the transformative aspirations of the NDP 2030.

The third pathway is restoring stability in local government through its relationship with society. The nucleus of South Africa's capable and developmental state is located in the stability of its local government sphere. Restoring stability in this sphere relies heavily on its relationship with society and establishing a multi-level stakeholder engagement approach towards achieving the country's national goal of a capable and developmental state. This means building trust on multiple levels and rebranding local government as an engine of social transformation open to working with society to enhance its capacity to create public value. In a context of minimal trust, we suggest multiple stakeholder engagements through the quadruple helix led by the state to re-emphasise the transformative intentions of NDP 2030 and the importance of service delivery for community well-being. A consensus generation model on service delivery actions predicated on the shared mutual goals of poverty reduction, employment generation, and inequality reduction must be prioritised. Underpinning this consensus model will be an understanding that South Africa is a developmental state that must provide services for a just and equitable society. This means that the best skill sets in the country need to be employed and supported to navigate the current service delivery failures in municipalities. It also entails that policy implementation is left to a skilled and capacitated human resource core in collaboration with sectors and stakeholders from society. Examples from East Asia show how a capacitated human resource base can be integral to shifting a country's economic fortunes and ensuring an improved quality of life for a just and equitable society.

The fourth pathway relates to the technical skills, financial management challenges, and labour relations issues confronting the municipal administration. This would require a rethink of the role of skills development in South Africa's local government sphere and how such skills are developed and supplied in relation to these areas. As part of the quadruple helix stakeholders, the higher education sector needs to be a key ally to the developmental agenda of the state. While higher education must meet the needs of the NDP 2030, there is a need to prioritise focus areas in line with current and projected future labour and developmental demands of the state. From a local government perspective, particularly specialised skills are required to understand the dynamic requirements of the municipal environment and its communities. Higher education must be complicit in producing skilled and trainable artisans and engineers who can maintain infrastructure and lead the service delivery revolution. This means creating a supportive policy and resource climate to increase the number of artisans and technicians in diverse engineering and science fields. Currently, there is an overemphasis on social science and administration at the Technical and Vocational Colleges (TVETs). Yet, the shortage of plumbers, boilermakers and other associated professionals has reached crisis levels. Adapting the German model of a strengthened TVET sector is important to ensure a capacitated human resource core to improve service delivery. Although there has been endless rhetoric on this, the main issue has been the poor implementation of the aspiration. TVETs have largely operated like very senior high schools and are not centres of innovation and invention.

Reflecting on 30 years of democracy and 26 years of developmental local government, there is a pressing need to revisit the state of local government in South Africa and its service delivery praxis. Looking back and forth, local government's stability and the extent to which it can stay resilient in a volatile, uncertain, complex and ambiguous service delivery environment is no longer vested in its own capacity, but that of society.

Declarations

We have no competing interests to declare. AI and LLMs were not used in the writing of this paper. Both authors read and approved the final manuscript.

References

1. Leydesdorff L. The triple helix, quadruple helix, …, and an N-tuple of helices: Explanatory models for analyzing the knowledge-based economy? J Knowl Econ. 2021;3(1):25–35. https://doi.org/10.1007/s13132-011-0049-41
2. Ukwandu C. South Africa as a developmental state: Is it a viable idea? Afr J Public Aff. 2019;11(2):41–62.
3. Bianchi C, Nasi G, Rivenbark WC. Implementing collaborative governance: Models, experiences, and challenges. Public Manag Rev. 2021;23(11):1851–1589. https://doi.org/10.1080/14719037.2021.1878777
4. Reddy PS. Evolving local government in post-conflict South Africa: Where to? Local Econ. 2018;33(7):710–725. https://doi.org/10.1177/0269094218809079
5. South Africa's 32 most 'dysfunctional' municipalities. BusinessTech. 11 January 2024 [cited 2024 May 20]. Available from: https://businesstech.co.za/news/trending/741577/south-africas-32-most-dysfunctional-municipalities/
6. Republic of South Africa. Constitutions Act 108 of 1996. Pretoria: Government Printers; 1996.
7. Van der Waldt G, Venter A, Phutiagae K, Nealer E, Khalo T, Vyas-Doorgapersad S. Municipal management serving the people. 3rd ed. Cape Town: Juta; 2018.
8. Department of Cooperative Governance and Traditional Affairs. State of Local Government in South Africa: National State of Local Government Assessment. Working document. Pretoria: Government Printers; 2009.
9. Department of Cooperative Governance and Traditional Affairs. Min Mec Briefing: The Local Government Turnaround Strategy. Pretoria: Government Printers; 2010.
10. Biljohn M. Social innovation and service delivery by local government [PhD thesis]. Bloemfontein: University of the Free State; 2018.

AUTHORS:
Gillian Maree[1]
Samkeliswe Khanyile[1]

AFFILIATION:
[1]Gauteng City-Region Observatory, Johannesburg, South Africa

CORRESPONDENCE TO:
Gillian Maree

EMAIL:
gillian.maree@gcro.ac.za

HOW TO CITE:
Maree G, Khanyile S. Moving beyond basic service delivery for inclusive reliable infrastructure. S Afr J Sci. 2024;120(11/12), Art. #18702. https://doi.org/10.17159/sajs.2024/18702

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
load shedding, water, income inequality, Gauteng, infrastructure

PUBLISHED:
30 August 2024

Moving beyond basic service delivery for inclusive reliable infrastructure

Significance:

In the last 30 years of democracy, the focus has been on delivering basic services to households. Yet there are larger infrastructure challenges faced at a local level, including rising dissatisfaction regarding the quality of services delivered. Insights generated from ongoing discussions on Gauteng's energy and water landscape indicate that, central to the service delivery discourse, delivering services goes beyond mere physical access. Using household survey data for Gauteng, we show it is necessary to make services more affordable and reliable, emphasising the need for innovation, inclusion, and placing equitable distribution at the core of infrastructural and service delivery planning.

Service delivery in Gauteng

South Africa celebrates 30 years of democracy in 2024, and while much progress has been made to address past inequalities and create opportunities for all citizens, many challenges remain. Issues of service delivery remain central to political debates now as much as they did in 1994. Over the past three decades, the government has focused on providing basic services to all citizens and reversing apartheid spatial planning. Underpinned by the South African Constitution, the first decade of post-apartheid South Africa saw a range of new legislation and policy development to guide the rollout of services. Significant progress has been made, and the 2022 Census showed that 88.5% of South Africans now live in formal dwellings, up from 65% in 1996.[1] A total of 82.4% of households have access to piped water in their homes or yards, and 94.7% have access to electricity. However, these statistics do not reflect whether those services are reliable or affordable. There is no denying that challenges remain to rolling out services to those households that still lack access, as evidenced by rising service delivery protests.

In Gauteng, significant progress has been made in the delivery of services. Over 90% of households have access to water, sanitation and electricity, and the recent national census has shown that access to those services has remained stable over time.[1] More so than in the rest of the country, housing in Gauteng is formalised, and the proportion of households living in informal dwellings has decreased from 20.2% (2011) to 11.5% (2022).[2] This trend also translates into service delivery. In Gauteng, the total number of households increased by 1 409 643 between 2011 and 2022, but 1 560 956 more households gained access to water inside their dwellings in the same period, reducing the service delivery gap. A similar trend exists for sanitation, where there has been a reduction in households that use a pit latrine (13.7% to 5.6%) and those that have no toilet facilities (2.4% to 0.4%). Access to electricity in Gauteng is high, but there are notable shifts in the choice of energy sources, most likely due to load shedding. These shifts include a significant change in the energy source for cooking, with a third of households (32%) using gas for cooking in 2022, compared with only 3.1% in 2011.

Increasingly, though, much of the conversation about infrastructure is about infrastructure failure, particularly driven by the impact of load shedding by Eskom. Increasing evidence supports the severe economic impact of electricity disruptions on the economy, households and society.[3,4] This is also true for water infrastructure, and Day Zero events in the Western and Eastern Cape, as well as rising water interruptions in Gauteng, which also impact households and the broader economy.[5,6]

Data on household satisfaction with water and electricity services and reported service interruptions can shed some light on the extent of the challenge in municipalities in Gauteng. Data show that reliability and satisfaction with basic services in Gauteng mirror the inequalities in South African society, with poorer and more marginalised communities struggling with the least reliable services. Table 1 presents data on reported household access to water and electricity services, service interruptions, and satisfaction with these services at a municipal level for 2020/2021. Generally, service delivery is higher in metropolitan areas, yet service interruptions are also high. The City of Johannesburg has the second-highest weekly water interruptions (12%) in Gauteng after Emfuleni (18%). This translates to 1 in every 10 households reporting weekly water supply interruptions. The numerous infrastructure challenges in Emfuleni municipality are evident, with reported satisfaction and service interruptions below the Gauteng average across all measures.

The City of Johannesburg provides a good illustration of the service delivery challenges in Gauteng. As an urban municipality with some of the highest levels of service delivery and infrastructure rollout in South Africa, some of the highest levels of water and electricity interruptions have been reported across Gauteng.

However, focusing only on service delivery misses some of the larger infrastructure challenges faced at a local level, with increasing dissatisfaction amongst citizens concerning the quality of services delivered. In the next section, we explore water and electricity services to understand the levels of service interruptions and their consequences, using Gauteng as a case study.

Service delivery case studies: Water and energy

Water interruptions

Increasingly, major cities around the world are experiencing conditions where urban growth and increasing demand for water have begun to stretch the limits of locally available water. Additionally, increasing climate challenges such as the unpredictability of rainfall and rising heat levels has led to a situation in which several cities have seen conditions where water demand exceeds supply, creating a water crisis. While the challenges in Cape Town may be

Table 1: Access, satisfaction and service interruptions for water and electricity* services expressed as the percentage of households in each municipality in Gauteng. Shaded areas indicate municipal values below the Gauteng mean.

Municipality	Piped water in the dwelling or yard	Always clean water	Access to adequate sanitation	Satisfied with sanitation services	Satisfied with access to water	Regular water interruptions*	Water interruptions every week	Use electricity for lighting	Use electricity for cooking	Regular electricity interruptions*	Electricity interruptions every week	Satisfied with energy sources	Satisfied with cost of municipal services
Ekurhuleni	93%	73%	96%	79%	89%	28%	6%	95%	86%	63%	20%	74%	38%
Johannesburg	95%	79%	96%	74%	87%	37%	12%	94%	84%	74%	30%	64%	32%
Tshwane	91%	72%	86%	69%	78%	33%	5%	93%	78%	66%	12%	69%	28%
Emfuleni	94%	73%	95%	59%	74%	49%	18%	88%	80%	74%	37%	56%	15%
Lesedi	94%	74%	95%	84%	92%	20%	2%	94%	76%	48%	10%	73%	29%
Midvaal	79%	79%	88%	73%	84%	16%	2%	83%	59%	53%	18%	66%	35%
Merafong	86%	73%	84%	51%	78%	38%	4%	92%	87%	69%	31%	65%	16%
Mogale City	87%	68%	90%	72%	84%	21%	5%	90%	78%	33%	9%	75%	31%
Rand West	86%	74%	83%	56%	80%	21%	4%	84%	76%	35%	9%	67%	20%
Gauteng (mean)	93%	75%	93%	73%	84%	33%	9%	93%	81%	64%	22%	68%	31%

Data source: Gauteng City-Region Observatory[7]. Public access data for use under a CC-BY licence.

**Regular water/electricity interruptions are defined as interruptions that occur at least once a month in a 12-month period.*

Table 2: Percentage of households reporting frequent water interruptions compared to income category in Gauteng

		Frequency of water interruptions over 12 months				
		Every week	A couple of times a month	Once a month	A couple of times a year	Never
Household income per month	ZAR1–800	12%	19%	10%	32%	27%
	ZAR801–3200	11%	18%	9%	34%	29%
	ZAR3201–12 800	9%	16%	9%	34%	31%
	ZAR12 801–25 600	5%	13%	8%	40%	34%
	ZAR25 601–51 200	4%	11%	6%	40%	39%
	ZAR51 201 or more	1%	8%	4%	46%	40%
Gauteng (mean)		9%	16%	9%	35%	31%

Data source: Gauteng City-Region Observatory[7]. Public access data for use under a CC-BY licence.

the most well known to South Africans, Mexico City, São Paulo, Lahore and Los Angeles have all recently experienced water challenges.

The most well-documented South African case is the Day Zero crisis in Cape Town in 2017/2018, when the city counted down the days of water supply it had left whilst also severely restricting water usage across the City. Gauteng also faced critical water shortages during a severe dry spell in 20152016. In the last two years, Gauteng has had further infrastructure and water supply challenges during periods of high demand. Gauteng's experiences with water shortages are a factor of water availability challenges, infrastructure challenges and urban dynamics.[8] In most cases, interventions into water are focused on water availability and infrastructural issues, and little consideration is given to residents' experiences with water or water infrastructure within urban growth and densification and how this enables or disables growth and development.

The challenge in Gauteng has become less about access to water infrastructure and more about the reliability and quality of services provided and consumption levels. Between 2018 and 2021, the number of households reporting frequent water interruptions (water interruptions once a month or more) rose by 3% to 33%, or a third of all households. Households in Johannesburg (37%), Merafong (38%) and Emfuleni (49%) were the most likely to report frequent water interruptions.[9]

Recent challenges in Gauteng with water supply have often resulted in water entities pointing to consumer overconsumption, particularly in areas like Alexandra, the Johannesburg CBD, and Soweto. Yet these typically high-density township areas are more likely to report water interruptions and have high levels of water leakages.[9] Townships and peripheral areas are more likely to experience challenges with both access and satisfaction with water.

Table 2 shows the frequency of water interruptions over 12 months (2020/2021) with household income. The lower the household income, the more likely these households are to experience water interruptions.

Comparing data on the access and reliability of water infrastructure shows that water delivery challenges mirror historical and current inequalities in South African society. Table 2 shows that, as incomes rise, so does the reliability of water services. For example, 12% of the poorest households (those earning less than ZAR800 per month) report weekly water interruptions compared to 1% of the wealthiest households (earning more than ZAR51 201 per month). Overall, the poorest households are the most likely to struggle with regular access to water services.

The data show that lower-income households are more likely to have negative experiences with water and that poor and more vulnerable households disproportionately bear the burden of water supply challenges. Although also affected, wealthier households are less likely to experience water supply challenges and more likely to have disposable income to shield themselves from service interruptions.

Electricity interruptions

Similarly, Gauteng has experienced electricity access and supply challenges, both as a result of load shedding and infrastructure failures at the local level. Over the past two years, Eskom's inability to meet the country's electricity demand has resulted in unprecedented load shedding, with the country experiencing 205 days of load shedding in 2022.[10] Electricity interruptions are affected by factors such as generation and distribution constraints, inadequate maintenance of ageing infrastructure leading to breakdowns, vandalism, lightning, transformer malfunctions, illegal connections, theft of electricity infrastructure, and debt.[11,12] Figure 1, from the GCRO's QoL Survey (2020/2021)[7], revealed that 94% of respondents had prepaid or postpaid electricity. Of these, 22% (or 1 in 5 households) experienced electricity interruptions at least once a week – a significant increase from 4.8% in the previous (2017/18) survey.[11] There has been a general increase in electricity interruptions across all municipalities in Gauteng. The most affected municipalities, with over 30% of respondents enduring weekly electricity interruptions, are Emfuleni (37%), Merafong (31%), and Johannesburg (30%).

Moreover, the spatial distribution of electricity interruptions is concentrated in areas in or near Wedela (West Rand District), Lawley (City of Johannesburg) and Tsakane (Ekurhuleni).[13] These areas are predominantly townships, state housing projects, and informal settlements characterised by rapid expansion and growth, high population densities, and infrastructure that was initially developed for much smaller populations.

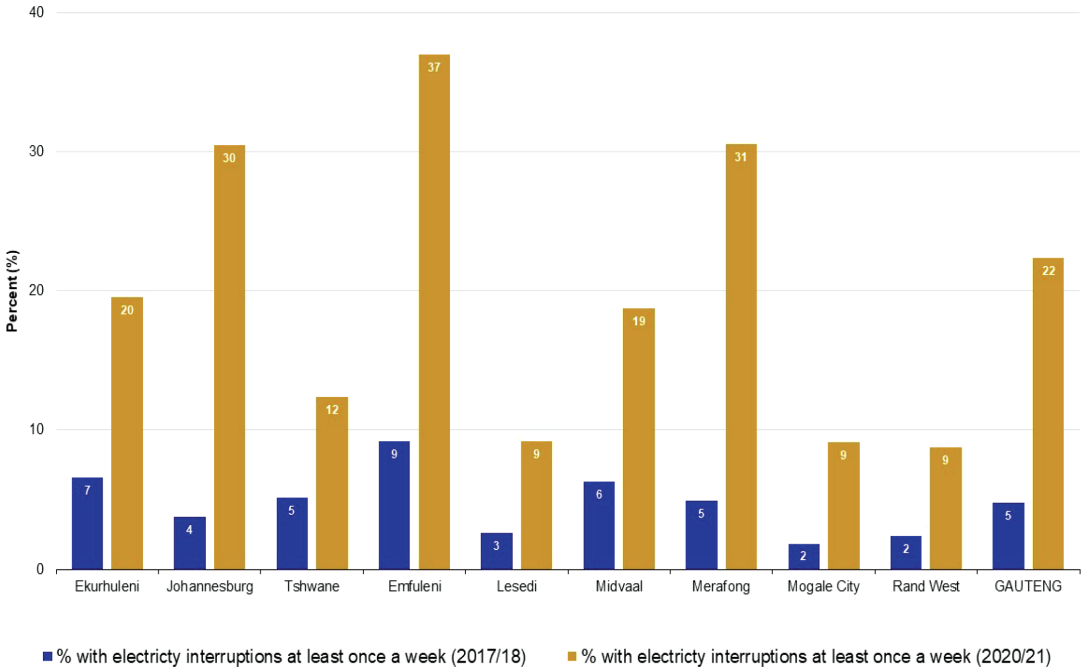
Following a similar trend to the water interruptions data, Table 3 further shows that households with lower incomes, such as 30% of those with a monthly household income between ZAR1 and ZA800, reportedly experience electricity interruptions weekly, compared to those in higher-income categories (earning ZAR12 801 or more per month).

These findings indicate that both water and electricity interruptions are more prevalent in predominantly lower-income, historically underserved communities. This is particularly concerning as the lack of access to electricity during these interruptions disrupts basic household activities and compounds existing challenges faced by underserved communities, such as a lack of adequate infrastructure and the delivery of other basic services. Moreover, lower-income households are less likely to have access to resources to shield themselves from the impact of load shedding. The unpredictable nature of electricity outages also impacts household food storage as food goes to waste. At the same time, electricity interruptions significantly impact informal and small businesses, the backbone of local economies, as they cannot operate during these times.[14]

Rising electricity interruptions are concerning, as access to electricity is critical for supporting economic development and growth. Ledger[15] argues that additional electricity generation capacity, of a third to half of the current capacity, is required to increase economic growth and development. In addition, electricity interruptions have also been noted to impact other infrastructure, such as water infrastructure. Specifically, in 2022, the severe electricity interruptions were accompanied by reports of damage to water infrastructure and supply challenges in Gauteng.[10,12]

Lessons from and consequences of electricity interruptions

Examining the transition to alternative electricity sources and going off-grid offers insights into the complexities and opportunities that define service delivery in Gauteng. Over the past seven years, the country has battled electricity generation and supply issues, resulting in a clear and concerning trend in the rise of load shedding days. This was followed by the South African government announcing incentives to encourage households and businesses to invest in alternative electricity sources in early 2023, in efforts to address the country's energy crisis. In this context, many residents in Gauteng have invested in alternative electricity sources to protect themselves from supply interruptions.



Source: Adapted from Mushongera et al.[13] with permission.

Figure 1: Percentage of respondents reporting electricity interruptions at least once a week in 2017/2018 and 2020/2021, by municipality.

More generally, the energy landscape in Gauteng and South Africa has changed in recent years, with more residents opting to use alternative energy sources. One of the notable aspects of this transition has been the dynamic nature of energy choices and adaptability. While individuals and communities seek alternatives to conventional grid-based systems, only about 1 in 20 households accessed alternative electricity in 2020/2021.[16] Although this is only a small minority of Gauteng residents, this number has been steadily increasing as load shedding increased, and the costs of the systems dropped. Only 5% and 4% of Gauteng respondents reported accessing solar or wind energy and diesel generators, respectively, by 2021.[16]

Data show that affluent households are more likely to invest in alternative electricity sources than poorer households.[12] Figure 2 shows that less than 5% of those households with an income below ZAR12 800 had access to solar and wind energy and generators. By contrast, 12% and 17% of those households with an income of ZAR51 201 or more had access to solar and wind energy and generators, respectively.

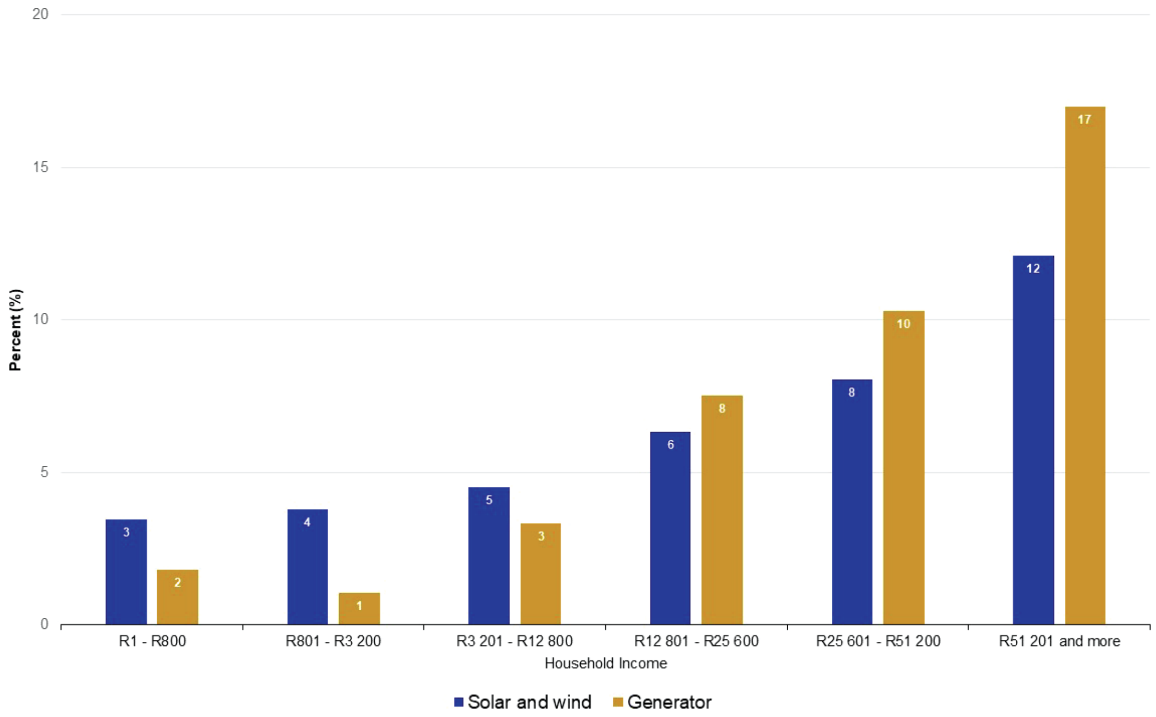
The transition to off-grid energy sources, especially among affluent households and businesses, has been critiqued as it could further widen the gap between affluent and poorer households and influence municipal service provision. Municipalities depend on revenue from basic services and utilities. Ledger and Rampedi[11] argue that moving to renewable energy sources may undermine municipalities' financial stability. However, municipalities could play a role in increasing energy generation capacity through renewable energy deployment.

The key drivers and consequences of going off-grid vary depending on socio-economic status, geographic location, and individual circumstances. Some households go off-grid to escape electricity interruptions and escalating tariffs. For instance, between 2007 and 2022, electricity tariffs increased by 653%, while inflation increased by 129% over the same period.[17] This means that the price of electricity increased by more than four times the inflation rate. Some households choose to go off-grid for sustainability and climate change reasons. Others do not choose to go off-grid but are instead forced into it by economic

Table 3: Percentage of households reporting frequent electricity interruptions compared to income category in Gauteng

		Frequency of electricity interruptions over 12 months				
		Every week	**A couple of times a month**	**Once a month**	**A couple of times a year**	**Never**
Household income per month	ZAR1–R800	30%	34%	9%	21%	6%
	ZAR801–3200	26%	35%	8%	25%	5%
	ZAR3201–12 800	23%	37%	8%	25%	6%
	ZAR12 801– R25 600	16%	37%	9%	30%	8%
	ZAR25 601–51 200	14%	32%	9%	37%	7%
	ZAR51 201 or more	14%	36%	11%	34%	5%
Gauteng (mean)		22%	36%	8%	27%	6%

Data source: Gauteng City-Region Observatory[7]. Public access data for use under a CC-BY licence.



Source: Adapted from Culwick Fatti and Khanyile[12] with permission.

Figure 2: Access to alternative electricity by income.

constraints or a lack of access to reliable services. Ledger[15] highlights the realities of the existing electricity tariff structure and its disproportionate burden on the economically disadvantaged, keeping them in a cycle of financial strain and energy insecurity. This can be seen in the communities caught in Johannesburg's electricity tariff trap, indicating how the system reinforces rather than lessens existing inequalities.

Parallel to these discussions on electricity pricing, the transition has been commended for its active role in climate change mitigation. However, it has also been critiqued for only offering temporary solutions to grid shortages, requiring further research to become more scalable and affordable[18], and leading to lower energy availability and higher associated costs for heating, cooking, and other needs compared to grid-connected homes.[19] Additionally, energy consumption rises particularly after electricity interruptions. All these factors add increasing complexity to the grid and grid maintenance.

Conclusion

The insights generated from ongoing discussions on Gauteng's energy landscape indicate that recognising that delivering services such as electricity goes beyond mere physical access to the service is central to the discourse on going off-grid. It is necessary to make services, such as electricity and water, more affordable and reliable, emphasising the need for innovation, inclusion, and placing equitable distribution at the core of infrastructural and service delivery planning.

The lessons from the transition to alternative sources of electricity indicate the far-reaching impacts of load shedding and should be a warning for municipal service delivery. Very little is understood of the longer-term consequences of residents attempting to take water and electricity services off-grid. Recent media reports indicate an increase in wealthier households installing backup water tankers or sinking boreholes to shield themselves from water interruptions. In December 2023, the City of eThekwini issued a statement urging residents to disconnect static tanks connected to the municipal water reticulation system in Phoenix, noting the impact on poorer households.[20] In this case, many wealthier households had installed static tanks to store water during prolonged water outages, placing pressure on the water reservoir to maintain adequate storage levels. Those poorer households without static tanks then faced even longer water outages as these static tanks were also filled.

Service delivery, particularly the delivery of basic services such as water and electricity, in Gauteng, has seen considerable strides since the end of apartheid. However, significant challenges still need to be addressed, particularly regarding the resilience and reliability of these services. Water and electricity interruptions disproportionately affect poorer communities, while affluent households are increasingly investing in alternative electricity and water sources in an attempt to go off-grid and bypass municipal services. Currently, some of these choices may be further entrenching existing inequalities in communities.

The challenges for service delivery for the next 30 years of democracy may well relate to the consequences of poor infrastructure maintenance and rising customer dissatisfaction.

Declarations

We have no competing interests to declare. We have no AI or LLM use to declare. Both authors read and approved the final manuscript.

References

1. Statistics South Africa (Stats SA). Census 2022. Pretoria: Stats SA; 2023.
2. Götz G, Ballard R, Hassen EK, Hamann C, Mahamuza P, Maree G, et al. Statistical surprises: Key results from Census 2022 for Gauteng [webpage on the Internet]. c2023 [cited 2024 Jul 05]. Available from: https://www.gcro.ac.za/outputs/policy-and-other-outputs/detail/statistical-surprises-key-results-from-census-2022-for-gauteng/
3. Akpeji KO, Olasoji AO, Gaunt C, Oyedokun DTO, Awodele KO, Folly KA. Economic impact of electricity supply interruptions in South Africa. SAIEE Afr Res J. 2020;111(2):73–87. https://doi.org/10.23919/SAIEE.2020.9099495
4. Wiese M, van der Westhuizen LM. Impact of planned power outages (load shedding) on consumers in developing countries: Evidence from South Africa. Energy Policy. 2024;187, Art. #114033. https://doi.org/10.1016/j.enpol.2024.114033
5. Matikinca P, Ziervogel G, Enqvist JP. Drought response impacts on household water use practices in Cape Town, South Africa. Water Policy. 2020;22(3):483–500. https://doi.org/10.2166/wp.2020.169
6. Loubser C, Chimbanga BM, Jacobs H. Intermittent water supply: A South African perspective. Water SA. 2021;47(1):1–9. https://doi.org/10.17159/wsa/2021.v47.i1.9440
7. Gauteng City-Region Observatory. Quality of life survey 2020-2021, round 6. Johannesburg: Gauteng City-Region Observatory; 2021. https://doi.org/10.25828/WEMZ-VF31
8. Muller M, Schreiner B, Vermeulen A, Maree GA, Reddy T. Water security perspective for the Gauteng City-Region: Securing water for continued growth and well-being [Internet]. Johannesburg: Gauteng City-Region Observatory; 2019. p. 72. Available from: https://cdn.gcro.ac.za/media/documents/GCR_Water_Security_Perspective_for_web_2019.pdf
9. Petersen C, Dlamini L, Maree G. Water interruptions in Gauteng [webpage on the Internet]. c2022 [cited 2024 May 19]. Available from: https://www.gcro.ac.za/outputs/map-of-the-month/detail/water-interruptions-gauteng/
10. Du Plessis A. Power cuts in South Africa are playing havoc with the country's water system. The Conversation. 2023 January 23 [cited 2024 May 19]. Available from:http://theconversation.com/power-cuts-in-south-africa-are-playing-havoc-with-the-countrys-water-system-197952/
11. Ledger T, Rampedi M. Hungry for electricity. Johannesburg: Public Affairs Research Institute; 2022.
12. Culwick Fatti CC, Khanyile S. South Africa's power crisis: going off the grid works for the wealthy – but could deepen injustice for the poor. The Conversation. 2023 February 22 [cited 19 May 2024]. Available from: http://theconversation.com/south-africas-power-crisis-going-off-the-grid-works-for-the-wealthy-but-could-deepen-injustice-for-the-poor-200288
13. Mushongera D, Nyuke S, Khanyile S. Electricity interruptions in the GCR [webpage on the Internet]. c2022 [cited 2024 May 19]. Available from: https://www.gcro.ac.za/outputs/map-of-the-month/detail/electricity-interruptions/
14. Olunloyo O, Moloi T. Power cuts are hurting small businesses in South Africa – but sharing resources and equipment might be a solution. The Conversation. 2023 March 15 [cited 2024 May 19]. Available from: http://theconversation.com/power-cuts-are-hurting-small-businesses-in-south-africa-but-sharing-resources-and-equipment-might-be-a-solution-200871
15. Ledger T. SA needs rapid municipal electricity supply diversification. Daily Maverick. 2024 April 29 [cited 2024 May 19]. Available from: https://www.dailymaverick.co.za/article/2024-04-29-rapid-municipal-energy-supply-diversification-will-light-way-to-a-much-brighter-economy/
16. De Kadt J, Hamann C, Mkhize SP, Parker A.Quality of life survey 6 (2020/21): Overview report. Johannesburg: Gauteng City-Region Observatory; 2021. https://doi.org/10.36634/2021.db.1
17. Moolman S. 2022 Update: Eskom tariff increases vs inflation since 1988 (with projections to 2024) [webpage on the Internet]. c2022 [cited 2024 May 19]. Available from: https://poweroptimal.com/2021-update-eskom-tariff-increases-vs-inflation-since-1988/
18. Tol RSJ. Navigating the energy trilemma during geopolitical and environmental crises [preprint]. arXiv; 2023 [cited 2024 May 19]. https://doi.org/10.56506/EBHD4081
19. Runsten S, Fuso Nerini F, Tait L. Energy provision in South African informal urban settlements – A multi-criteria sustainability analysis. Energy Strategy Rev. 2018;19:76–84. https://doi.org/10.1016/j.esr.2017.12.004
20. eThekwini Municipality. Media release: City urges residents to disconnect static tanks connected to municipal reticulation system to conserve water [webpage on the Internet]. c2023 [cited 2024 May 17]. Available from: https://www.durban.gov.za/press-statement/City+Urges+Residents+to+Disconnect+Static+Tanks+Connected+To+Municipal+Reticulation+System+To+Conserve+Water/

AUTHOR:
Craig Sheridan[1,2]

AFFILIATIONS:
[1]Claude Leon Foundation Chair in Water Research, University of the Witwatersrand, Johannesburg, South Africa
[2]Director: Centre in Water Research and Development, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Craig Sheridan

EMAIL:
Craig.Sheridan@wits.ac.za

HOW TO CITE:
Sheridan C. Water insecurity in Johannesburg (and beyond?). S Afr J Sci. 2024;120(11/12), Art. #19068. https://doi.org/10.17159/sajs.2024/19068

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
Johannesburg, water security, reclamation, intermittent water supply

PUBLISHED:
30 August 2024

Water insecurity in Johannesburg (and beyond?)

Significance:

This Commentary draws upon news and other published information primarily from the last few years. The purpose of the article is to demonstrate that the water sector is complex, and that it sits at the interface of politics, science, engineering, social and behavioural science, economics, and public health. The lack of attention paid to maintenance and upgrades in the water sector, whilst demographic shifts occur, has resulted in water insecurity in South Africa's most populous province – Gauteng. The long-term prognosis for this situation is not good, unless we can mobilise society towards changing its understanding of how to fully reuse water and embrace the reuse of fully treated sewage. This represents a challenge that may take some time and will certainly require all the disciplines (and more) listed above.

Introduction

South Africans are familiar with load shedding. It has been a part of life since 2007 when we first experienced rolling blackouts as a result of insufficient capacity of the electricity supply grid. South Africans are also quite resilient; and over the last two decades, we have come up with many strategies to cope with an intermittent electric power supply. These strategies have included various battery backup systems, various solar installation systems, and a major switch from electric stoves to gas hobs and from electric geysers to gas or solar-thermal geysers.

This experience of load shedding has become the normal way of life for most of us. Unfortunately, 'shedding' has moved beyond electric power and firmly into the domain of water supply. Gauteng, and particularly the Johannesburg Metro, has started to be subjected to significant occurrences of 'water shedding'[1] or intermittent water supply (IWS).

South Africa has engineered effective systems to reduce and prevent water insecurity. Systems of dams and inter-basin transfer schemes have been built[2] across southern Africa to continue supplying water to primarily Johannesburg (and Gauteng), the industrial heartland of the country and the centre of South Africa's population mass[3]. This strategy was highly effective at ensuring water security, until very recently.

IWS is not only a South African phenomenon. Approximately one billion people globally have water connections which regularly undergo IWS.[4] The regular occurrence of IWS in Johannesburg is, however, a new phenomenon. IWS as a strategy for managing water supply has serious negative consequences, including increasing residential (i.e. household) capital outlay for water storage through the need for additional capacity, such as JoJo tanks and booster pumps[4], causing damage to the distribution network[5], and compromising the quality of the water, particularly with regard to its microbiological profile[6-8].

IWS in Gauteng – is this a scarcity or security issue?

Water scarcity is defined as an excess of water demand over available supply. This is generally signalled by an unsatisfied demand or tensions between users or competition for water or overextraction of groundwater and insufficient flows to the natural environment.[9] The definition of water security, according to United Nations Water, is[10]:

> *The capacity of a population to safeguard sustainable access to adequate quantities of acceptable quality water for sustaining livelihoods, human well-being and socioeconomic development, for ensuring protection against waterborne pollution and water related disasters and for preserving ecosystems in a climate of peace and political stability.*

In South Africa, water security is ensured by the Department of Water and Sanitation (DWS). They allocate water based on the population of a region and the availability of water in the storage systems (i.e. dams). This allocation is sold by the DWS to the bulk water utility (Rand Water for Gauteng, including Johannesburg) which treats it to obtain a potable quality. Rand Water sells this fresh potable water to a number of metropolitan municipalities (for Johannesburg the metro entity is Joburg Water), which then sells it onwards to consumers. There is currently, however, a mismatch between what the DWS sells and what is needed, as the DWS is responsible for supplying both current and future needs. South Africa, and particularly Johannesburg, is not currently water scarce (although this could change quite rapidly if we were to experience a day-zero type drought), but is currently very water insecure.

Johannesburg is located on the African continental divide at an altitude of approximately 1600 metres above mean sea level and has no large rivers or natural water sources. As a consequence of its geographical location, all its water is imported. To continue to maintain water security, additional dams are being built in Lesotho as part of the Lesotho Highlands Transfer Scheme. Unfortunately, the construction of these dams has been delayed by 8 years. This delay coincides exactly with a period in which Gauteng grew from 12 million to 15 million people. This means that in 2023, the province had the same water storage for a population which had grown by 25%.

Over the last 3–4 years, the City of Johannesburg has increasingly supplied water intermittently to its customers. In the summer of 2023/2024, water outages became a regular occurrence, especially for those living at higher elevations or in older suburbs in the city. These characteristics fit all the criteria for water insecurity, i.e. an excess of demand over supply.

In the older suburbs, the initial cause of these outages was primarily due to leaking water supply pipes which were patched (note – good engineering practice would define this as a *temporary* repair). During the patching process, water to the entire suburb would be shut off. This temporary nature of intermittent supply has now totally

changed, and this is indicative of a larger system which is not being properly maintained. When some minimal maintenance is conducted, the job is not properly completed. As a result, the roads are left with large, open excavation works, sometimes for months on end following a water repair, and the water repair often still leaks.[11] This is happening across all suburbs now.

Whilst this was (and is) happening in Johannesburg, in Hammanskraal in Pretoria there was an outbreak of cholera in 2023 which claimed over 30 lives.[12] At the same time at which both of these events were occurring, the blue[13] and green[14] drop audits and reports were re-established and published. Both reports indicate that the freshwater treatment works supplying our drinking water, as well as the wastewater treatment works treating our sewage, are *increasingly dysfunctional*. This trend is seen across the entire country. In addition, a no-drop report[14] was released which shows the amount of water which is stolen or lost through leaking pipes. For Johannesburg, this amount is close to 50%. This means that the allocation of water for Gauteng which is distributed to Johannesburg, Pretoria and other cities, is reduced by 50% (in the case of Johannesburg), because of theft and/or leakages. The DWS will not increase the quota for this province because they are the stewards or the custodians of the resource at a national level. They are concerned with meeting both current and future needs for the entire country, not just the immediate needs of Gauteng, or Johannesburg.

These statistics indicate that we have 50% less water for 25% more people. This extra usage and loss overloads the system entirely. The rate of drawdown on our potable water reservoirs is greater than the maximum rate of recharge. This discrepancy has resulted in reservoirs running empty and the City of Joburg throttling back supply during periods of high demand in order for reservoirs to recover, leaving many residents without water, not just for nightly periods but in some instances for weeks.[15] For those in the highest-lying areas (i.e. those living higher up the hill) the impact is the greatest. In addition to this, the occurrence of load shedding and power supply challenges means that pumping stations are often also not working. This is the current state of water systems, especially in Johannesburg. The blame for this poor state must be firmly placed with the Johannesburg City Administration.

There is, however, one critical difference between water and electricity. It is possible to live for weeks without electricity, as many in South Africa do. It may not be a comfortable lifestyle; however, it is a possible lifestyle. Without water, the options rapidly diminish in terms of for how long people can be resilient. The cost of purchasing bottled/container water is prohibitive, and yet, under these circumstances, it becomes the only available option. For those who grow food in their gardens (as happens in many poorer settlements), the cost increases exponentially because the allotment-type gardens which contribute towards food security are dry and barren from lack of water. The poorest have to purchase not only water but additional food too.

The basic domestic usage of water is not for swimming pools or gardens, but rather for sanitation (flushing toilets, washing hands, bathing), cleaning vegetables, cooking food, and drinking. These activities require a steady supply of water, even if only a little, and for this reason, the state provides about 6 kL per household per month free of charge. Therefore, the ongoing water outages place an incredible strain on a society which, whilst incredibly resilient, is wasting that resilience on dealing with challenges which it should not actually be dealing with. That resilience could be better utilised by seeking meaningful employment or furthering education, rather than by thinking about ensuring a reliable source of water and the increased cost of food. Obviously, these consequences affect the poor far more than the rich.

Long-term prospects

What are the long-term prospects for a resolution to this problem? Unfortunately, there is very little space left in South Africa to construct new large dams. This means that water security will decline after 2028, especially if we proceed with current social, cultural, political and engineering practices. Additionally, there is the confounding impact of the climate crisis: this drives up water demand as temperatures are higher, it increases evaporative losses on the dams and there is an increased number of flood events from bigger storms which pose the very real risk of destroying infrastructure such as dams, water treatment plants and wastewater treatment plants (such as occurred in Durban in 2022[16]). Indeed, the UN Food and Agriculture Organization reported that South Africa's water stress increased from 61% in 2016 to 65% in 2020.[17] Johannesburg, in particular, notwithstanding the high rates of water lost from broken infrastructure, still uses too much water per capita – more than any other province.

Changes need to be made to ensure long-term sustainability and to improve water security. First, real shifts in the economics and pricing of water must occur. The value of water, especially potable water, is ZAR25 per 5 L of treated water purchased at a supermarket when there is no water available in a tap. At the higher tiers of municipal billing, the cost is only ZAR17/kL. Much steeper tariffs can be applied to users of water that exceed their fair share (although defining a fair share is somewhat contentious) and this will undoubtedly drive consumption downwards. The administrations of all cities in South Africa, particularly Johannesburg, need to urgently ring-fence funding for maintenance and infrastructural improvement.

As a global society, we need to reconsider how we value water. Our future requires some truly creative problem-solving. We might need to consider how to bring water from the African tropics (such as transfers from the Congo belt or the Zambezi). Once we have solved our energy crisis, the opportunity for desalination of sea or mine water for augmentation of potable supply might become a viable possibility. The future, however, must include direct reclamation (sewage to potable water) as the global population moves towards 10 billion people. In eThekwini (Durban), reclamation has been used to supplement industrial supply[18], and in Windhoek, Namibia, reclamation has been a way of life since 1968[19].

There is, however, still a critical need to conduct additional research to understand how pathogenic organisms (including emerging pathogens and viruses) and emerging contaminants (pharmaceuticals, per- and polyfluoroalkyl substances (PFAS), pesticides etc.) behave in wastewater treatment plants, such that when we return the treated sewage, especially to the potable water reticulation systems, we do not harm our societies with increased exposure to these compounds at potentially harmful levels (these will build up, possibly to toxic levels, if they are not removed at wastewater treatment plants). Significant civic education needs to take place to remove the 'yuck' factor of drinking treated sewage, despite the fact that we already drink water sourced from our rivers that are highly contaminated by sewage and wastewater treatment plant outfall, made all the worse by our failing wastewater treatment works. Society, including the water sector, needs to embrace true circular thinking: if we are good engineers, we can derive significant value from our wastewater, and, if we do this properly, our new waste stream will be clean potable water. The question we need to ask is: are we ready for this revolution?

Declarations

I have no competing interests to declare. AI was not used in the writing or editing of this manuscript.

References

1. Ntuli B. Water shedding is our next big challenge, but can it be prevented if we tackle the causes now. Daily Maverick. 2023 May 02 [cited 2024 Apr 24]. Available from: https://www.dailymaverick.co.za/article/2023-05-02-water-shedding-is-our-next-big-challenge-but-it-can-be-prevented-if-we-tackle-the-causes-now/
2. Haasbroek B. Orange-Senqu infrastructure catalogue: Overview [webpage on the Internet]. c2013 [cited 2024 Apr 24]. Available from: https://wis.orasecom.org/orange-senqu-infrastructure-catalogue-overview/
3. Statistics South Africa. Census 2022 [webpage on the Internet]. c2023 [cited 2024 Apr 24]. Available from: https://census.statssa.gov.za/assets/documents/2022/P03014_Census_2022_Statistical_Release.pdf
4. Meyer DDJ, Singh S, Singh J, Kumar M, He M. Learning from intermittent water supply schedules: Visualizing equality, equity, and hydraulic capacity in Bengaluru and Delhi, India. Sci Total Environ. 2023;892, Art. #164393. https://doi.org/10.1016/j.scitotenv.2023.164393

5. Christodoulou SE, Fragiadakis M, Agathokleous A, Xanthos S. Vulnerability assessment of water distribution networks under abnormal operating conditions and nonseismic loads – the case of intermittent water supply (IWS). In: Christodoulou S, Fragiadakis M, Agathokleous A, Xanthos S, editors. Urban water distribution networks: Assessing systems vulnerabilities and risks. Oxford: Butterworth-Heinemann; 2018. p. 131–159. https://doi.org/10.1016/B978-0-12-813652-2.00003-7

6. Bivins A, Lowry S, Wankhede S, Hajare R, Murphy HM, Borchardt M, et al. Microbial water quality improvement associated with transitioning from intermittent to continuous water supply in Nagpur, India. Water Res. 2021;201, Art. #117301. https://doi.org/10.1016/j.watres.2021.117301

7. Calero Preciado C, Husband S, Boxall J, del Olmo G, Soria-Carrasco V, Maeng SK, et al. Intermittent water supply impacts on distribution system biofilms and water quality. Water Res. 2021;201, Art. #117372. https://doi.org/10.1016/j.watres.2021.117372

8. Bautista-de los Santos QM, Chavarria KA, Nelson KL. Understanding the impacts of intermittent supply on the drinking water microbiome. Curr Opin Biotechnol. 2019;57:167–174. https://doi.org/10.1016/j.copbio.2019.04.003

9. UN Food and Agriculture Organization. WASAG – The global framework on water scarcity in agriculture [webpage on the Internet]. No date [cited 2024 Apr 24]. Available from: https://www.fao.org/wasag/en/

10. United Nations. Water security and the global water agenda [webpage on the Internet]. c2013 [cited 2024 Apr 24]. Available from: https://www.unwater.org/publications/water-security-and-global-water-agenda

11. Carnie T. Billions down the SA big city revenue drain as 40% of purified water is lost to pipe leaks [webpage on the Internet]. c2024 [cited 2024 Apr 24]. Available from: https://www.dailymaverick.co.za/article/2024-04-06-billions-down-the-sa-big-city-revenue-drain-as-40-of-purified-water-is-lost-to-pipe-leaks/

12. Evans J. Hammanskraal cholera outbreak 'represents the ears of the hippopotamus' of SA's wastewater treatment crisis [webpage on the Internet]. c2023 [cited 2024 Apr 24]. Available from: https://www.dailymaverick.co.za/article/2023-07-26-hammanskraal-cholera-outbreak-represents-the-ears-of-the-hippopotamus-of-sas-wastewater-treatment-crisis

13. Department of Water and Sanitation. Blue drop report: National 2023 [webpage on the Internet]. c2023 [cited 2024 Apr 24]. Available from: https://ws.dws.gov.za/IRIS/releases/BDN_2023_Report.pdf

14. South African Department of Water and Sanitation. Latest reports [webpage on the Internet]. c2023 [cited 2024 Apr 24]. Available from: https://ws.dws.gov.za/IRIS/latestresults.aspx

15. Banda M. Make a plan and fix Gauteng's massive water outages, activists tell Joburg City and Rand Water [webpage on the Internet]. c2023 [cited 2024 Apr 24]. Available from: https://www.dailymaverick.co.za/article/2023-02-27-make-a-plan-and-fix-joburg-water-outages-activists-tell-utilities/

16. Magubane T. State of eThekwini rivers dire after sewage spills [webpage on the Internet]. c2022 [cited 2024 Apr 24]. Available from: https://www.iol.co.za/mercury/news/state-of-ethekwini-rivers-dire-after-sewage-spills-ad37bebb-04cd-434e-9522-370949ccf45c

17. UN Food and Agriculture Organization. AQUASTAT – FAO's global information system on water and agriculture [webpage on the Internet]. No date [cited 2024 Apr 24]. Available from: https://www.fao.org/aquastat/en/

18. Pillay SD, Friedrich E, Buckley CA. Life cycle assessment of an industrial water recycling plant. Water Sci Technol. 2002;46(9):55–62. https://doi.org/10.2166/wst.2002.0204

19. Haarhoff J, van der Merwe B. Twenty-five years of wastewater reclamation in Windhoek, Namibia. Water Sci Technol. 1996;33(10):25–35. https://doi.org/10.2166/wst.1996.0658

AUTHORS:
Marc Kalina[1,2]
Catherina Schenck[3]

AFFILIATIONS:
[1]Department of Mechanical and Process Engineering, ETH Zürich, Zürich, Switzerland
[2]School of Engineering, University of KwaZulu-Natal, Durban, South Africa
[3]DSI/NRF/CSIR Chair in Waste and Society, Centre for Interdisciplinary Studies of Children, Families and Society, University of the Western Cape, Cape Town, South Africa

CORRESPONDENCE TO:
Marc Kalina

EMAIL:
marc.kalina@gmail.com

HOW TO CITE:
Kalina M, Schenck C. What a mess: Rethinking municipal waste management 30 years into South African democracy. S Afr J Sci. 2024;120(11/12), Art. #18995. https://doi.org/10.17159/sajs.2024/18995

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
solid waste management, governance, South Africa, inequality, service delivery

PUBLISHED:
30 August 2024

What a mess: Rethinking municipal waste management 30 years into South African democracy

Significance:

Within South Africa, local governments hold the mandate for providing waste management services. Unfortunately, 30 years into our democracy, reliable waste management services remain out of reach for vast segments of the population – a situation which contributes to environmental degradation and increased inequality. Safeguarding service delivery and addressing inequalities requires deep structural changes and a rethinking of our waste management systems. This must include the decentralisation of waste management services to incorporate all stakeholders within the waste value chain, the depoliticisation of local government, and reconsideration of the financial model to allow for a basic level of services to all.

Introduction

In March 2024, members of the South African Municipal Workers' Union (SAMWU) engaged in a period of strike action in eThekwini Metropolitan Municipality that saw severe disruptions to basic municipal services, including refuse collection. In the absence of waste management services, uncollected refuse littered the street, with piles of black bags stacked outside homes and loose pieces of rubbish blown about by the wind and scattered from torn bags, choking pavements and green space (Figure 1). Durban's streets had never been dirtier. No community was unaffected, with rich and poor neighbourhoods alike experiencing service disruption. Yet, although all experienced the strike, its impact was not felt equally. Affluent communities were able to minimise the impact through privatised cleaning and collection services and the ability to bring uncollected refuse back inside (facilitated by the understated luxury of space). However, within eThekwini's informal communities and townships, where municipal waste collection was already irregular or inadequate, uncollected refuse contributed to ever-growing informal dumping grounds. Already a common reality within poor communities, these dumpsites expanded and then overflowed during the strike, washing untold tons of waste into rivers and streams, and transforming vibrant communities into squalid and unsanitary slums.

Although the strike was eventually called off and municipal workers returned to start clearing out the collection backlog, impacts of the strike have lingered, even months later. This is audible in the growing scepticism expressed by ratepayers across the municipality, where communities are reluctant to return to a status quo where disruptions within the municipality can cause such havoc on their streets. Yet, the impacts are even more tangible in low-income communities and informal settlements like Johanna Road, which remain filthy and where dumpsites have only grown since March 2024.

Within South Africa, municipal governments hold the mandate for providing waste management services to all communities.[1] However, strike action aside, South Africa's municipalities have become increasingly unreliable service providers. Plagued with a decreasing taxpayer base relative to the population and increasingly fragile municipal governance and financial management, the financial integrity of South Africa's municipalities has been in

Figure 1: An unemptied bin overflows in Durban's Bulwer Park.

an alarming decline.[2] Furthermore, although South African municipalities have made major strides in addressing waste management service delivery gaps since the start of democracy and extending solid waste management services to many previously underserviced communities, inequality still underpins access to waste management systems across the country, determining who can or cannot access or provide sustainable services. For the millions who live within urban informal settlements or within the countless rural communities stretched across the country, *any* municipal waste management services, let alone reliable or regular services, remain a distant prospect.[3] As the expansion of sufficient municipal waste services into underserviced communities remains unlikely within the near future, and the increased fragility of municipal finances threatens to disrupt service delivery within communities where it is already established, we must ask: is this a system we can continue to rely on, and, more importantly, did it ever work for the majority to begin with? Or is it perhaps time to consider alternatives to the centralised waste management services that we have come to expect as *de rigueur* in South Africa's urban areas, and to explore new solutions for communities where municipal services have always been out of reach?

Municipalities and waste management

As noted, within South Africa, local municipalities hold the mandate for providing waste management services to communities. At the national level, the Department of Forestry, Fisheries and the Environment (DFFE) is responsible for the drafting of legislation and regulations that set national norms and standards for waste management, while the respective provincial departments are the responsible licensing authorities for waste management activities and facilities, such as landfills. However, it is the municipality that is ultimately responsible for implementing waste management service delivery.[1] Financed primarily through local tariffs for waste services (such as landfill gate fees, which are also set by each municipality), and supplemented by allocations from property rates, it is local and metropolitan municipalities that provide waste management services, such as waste collection, waste disposal, street cleaning and any waste minimisation or recycling efforts.[1] District municipalities hold very little of this mandate, but are responsible for bulk infrastructure that may be used by more than one local municipality.

Unfortunately, South Africa's municipalities have increasingly struggled to deliver on their service delivery mandate. This is particularly true within local municipalities, which are home to almost 40% of the country's population, where a combination of compounding factors ranging from urban migration, skills shortages and, most importantly, poor financial management have had serious impacts on municipal capacity and financial integrity.[4,5] Although observers have been drawing attention to these trends for more than two decades, and municipal integrity has been a challenge since the official start of democracy in 1994, the deterioration in municipal stability seems to have accelerated since the COVID-19 pandemic.[2] For instance, for the 2021/2022 financial year, only 38 out of 257 municipalities achieved clean audits – a further decline from the already abysmal 41 clean audits the previous year.[6] Although this is not as low as the figure in 2010/2011, when only five municipalities received clean audits, it represents a reversal from modest improvements shown in the late 2010s, which peaked at 48 clean audits achieved in the 2016/2017 financial year. Furthermore, as of 2023, nearly two-thirds of South Africa's municipalities are under financial stress, and, despite efforts by the National Treasury to assist embattled municipalities in the immediate post-COVID-19 period, it is expected that the financial situation of most beleaguered municipalities is likely to continue to deteriorate.[7]

The consequences of this steady deterioration in municipal finances have been increasingly dysfunctional municipalities, manifesting in crumbling municipal infrastructure, administrative instability, declining standards of living and, notably, increased service delivery failures.[7] Although the full implications of the disintegration of the South African state at the municipal level for waste management service provision, in particular, and service delivery more broadly, have yet to be fully understood, some consequences have begun to manifest clearly. For instance, internationally, the increased hollowing-out of public service provision has been observed to degrade the quality of public services[8] and to work to the exclusion of those unable to pay, restricting access for low-income individuals[9]. These trends have also been observed in a broad swath of distressed municipalities across the country. For instance, in Makana in the Eastern Cape, municipal collapse contributed to privatised collection for those who could afford to pay, and no collection for those who could not.[2] It has also manifested through persistent illegal dumping and non-enforcement of bylaws in rural municipalities such as Maruleng in Mpumalanga and Hantam in the Northern Cape.[3,10,11] Although these examples remain exceptions, trends suggest that they may be an indication of what is in store for all South Africans. If South African municipalities continue to retreat from their service delivery mandate, what are the consequences for South Africa's citizens who rely upon municipal systems daily? Furthermore, as an increasing number of citizens are able to turn away from the state – either through privatisation, migration or non-payment – what are the implications for municipal systems that rely on rates and fees to keep services running? Finally, for those who have already slipped through the cracks of municipal service, or for whom regular waste collection has always been a luxury, what are the consequences of increasingly inaccessible services and increasingly exclusive alternatives?

Waste management inequalities

Inequality determines access to waste management systems across the globe, defining who can or cannot provide or access services.[12,13] This is particularly true within South Africa, where inequality is deeply rooted in immense historical injustices and continues to leave an indelible (and highly visible) stain on our society. South African municipalities are characterised by inequality, and so are the services they provide. For instance, nearly half of the country's population continues to lack access to waste collection services, predominantly in non-white communities and historically marginalised rural and peri-urban areas.[11,14,15] Even within better-served metropolitan municipalities, vast sections of our urban populations are denied access to adequate or even basic waste management services, based on either their inability (or presumed inability) to pay property rates, or their lack of title to their land.[16] This inequality has been well documented within township settings, as well as within informal settlements[2,17], which include the growing (and precarious) communities often founded illegally on municipal land[18-20], and, as in the context of Cape Town's 'backyarders', informal structures erected on the property of formal homes – often in the backyard[10].

As mentioned, since 1994, the South African state has made major strides in addressing service delivery gaps rooted in historical inequalities. Yet, despite this progress, it is evident that municipal waste management structures have never really worked for vast sections of the population and, despite 30 years of democracy, have largely failed to provide adequate, relevant or appropriate services to rural or low-income communities. Although those of us within formal communities who benefit from regular waste collection may still decry the deterioration of municipal services, the loss is felt most severely in the country's declining prospects for addressing this inequality. There is growing evidence that the retreat of the state in waste management service provision will drive inequality, and, although we lack a clear understanding of how these inequalities will manifest and how severe their impacts will be, a few points have become apparent.

First, disruptions in waste management services are an inconvenience for the affluent, but for the poor it can be catastrophic, endangering public health and threatening our goal of creating healthy human settlements. This difference was clearly discernible within Durban's 2024 strike action, as well as earlier in Makana, which saw Makhanda's townships transform into "slums" and "dumpsites" as a consequence of the suspension of municipal waste collection services.[2]

Second, as public services deteriorate, the wealthy are largely able to supplement gaps in coverage with private service providers, while the poor get left further behind. For waste management, this has largely manifested in private collection and cleansing services, either individually arranged or coordinated through private housing estates, Urban Improvement Precincts, City Improvement Districts, or other ratepayer organisations. Although eThekwini may have suspended official collection and cleansing

services during the strike action, in affluent areas like uMhlanga Rocks, street-cleaning activities continued as usual and residents in gated housing estates would have seen their refuse bags collected like normal.

Although this creeping privatisation of waste services has shown clear benefits for those who *can* pay, it raises a number of serious concerns. First, privatisation globally has been found to lead to poorer quality services overall, and especially so for the poor who *cannot* afford to pay and are often excluded from services altogether.[8,9] Within South Africa, neoliberal approaches to public service provision have been observed to work to the detriment of the poor who cannot afford privatised services, compounding inequality in services received.[21,22] Furthermore, the ongoing commodification of waste in metros like Tshwane, Johannesburg and Cape Town, has encouraged private-sector participation in recycling and circular economy efforts, but has worked to exclude the poorest from participating within waste and resources markets.[23,24] Finally, although the private sector may fill gaps in service provision in a useful way, once municipalities cede their mandate, that mandate can be tough to claw back, with Santos[25] arguing that privatisation is akin to a one-way street and that the 'restatisation' of services can be a challenge once the private sector has been admitted.

Therefore, although the private sector certainly has a large role in waste management service delivery, and many South African citizens are justifiably wary of their local municipalities, the private sector alone will not produce a more equitable service delivery landscape. If the state is unable to provide services, and the private sector is unwilling to provide services to those who cannot pay, then what alternatives remain? If we tie waste management service provision to an individual's ability to pay, we must accept the reality that the majority of South Africans cannot afford to pay for waste management services, and therefore will not be able to access them.[26] Moreover, if we continue to deny individuals and communities access to basic waste management services, then we must accept the inevitability of continued environmental degradation, as well as the erosion of our democratic project that goes hand in hand with denying citizens basic human rights and dignity.

Exploring alternatives

Durban is not an outlier, and, although it took strike action and severe disruptions to service delivery to drive it home to the most privileged, the perilous state of municipal waste management in South Africa should concern us all. Yet, given the fragility of South African municipalities, and their own historical shortcomings at addressing entrenched inequalities, what is the most sustainable path forward, and what alternatives remain if we can no longer rely on the state?

Ultimately, as we have argued before[2], a dramatic improvement in municipal governance is likely the smoothest pathway towards restored – and hopefully improved and expanded – service delivery. However, at the time of writing, the likelihood of a sudden and significant improvement in municipal health seems low, and although national government has increasingly demonstrated concern about poorly performing municipalities, the track record of national intervention (and broader concerns about governance at the provincial and national level) suggest that this is not a silver bullet for turning around failing municipalities.

Without intervention, the proliferation of neoliberal alternatives, including privatised collection, processing and disposal, are the most likely outcome. Yet, this is a future we must be extremely wary of, as the private sector is inherently profit-seeking and will only service those who can pay, and South Africa's own experiences of privatisation in the waste sector suggest increased inequality, and often poorer services.[2,21,22] What other alternatives exist? Although we do not offer a complete roadmap, we propose a number of radical solutions that should be considered, given the current status quo.

First, we recommend a total reconceptualisation of the municipal system, including the depoliticisation of local government in order to focus on their core mandate of service delivery.[27,28] Although potential replacements could take numerous forms, professional and functional service delivery hubs, with monitoring and evaluation structures to ensure accountability from the state, residents and civil society, are a must. Examples relevant to South Africa can be found in Wilson et al.[28], who refer to participatory community-based waste management in countries such as Brazil, Mali, the Philippines and various sub-Saharan African countries, including Kenya and Tanzania. These locally developed service delivery models include NGOs, the informal sector, women's groups and local youth to provide services for all.[28]

Second, if we persist in providing adequate services only to those who can pay or can be linked to a municipal account, then we must accept the accompanying environmental degradation that accompanies denying a vast section of our population the right to a clean and healthy environment.[16] According to the 2022 Census, South Africa has 4297 informal areas housing over 2 million households, in addition to the multitude of backyarders in townships, which represent South Africa's fastest-growing housing sector.[29] Not rendering services to these households not only denies citizens a basic right, but affects all communities through the degradation of water sources, land, animal habitats and the local environment. Johanna Road's dumpsites of today, which straddle the uMngeni River, will become Durban's beachfront rubbish of tomorrow. Rather, we propose that the financial model for waste management services must be reconsidered, moving away from providing services only to those who pay rates and can be linked to a municipal account to providing a basic level of services to all, regardless of payment.[16] To ensure financial sustainability, studies such as those by Lohri et al.[30], Mokgabodi[31] and Coffey and Coad[32] recommend the use of more than one revenue stream such as rates and taxes, and national government grants (Municipal Infrastructure Grants and Urban Settlement Grants[31,32], and cross-subsidising from other income streams (water, electricity, sanitation), extended producer responsibility or polluter-pay principle fees, recycling and selling, and public-private initiatives. Local solutions should also be explored to develop cost-effective and locally relevant services. Godfrey et al.[33] argue for the strengthening of the implementation of policies and good governance to attract investments in waste management infrastructure and services.

Finally, to deliver efficient and appropriate waste management services, we believe that existing service arrangements must be reconfigured and co-designed with local communities. Core members of new waste management plans must include all role players in the waste value chain, including informal waste pickers, private-sector actors and civil society. Furthermore, the strike in Durban has emphasised that informal waste pickers *work*, even when the municipality does not. Although there was no formal collection, tons of recyclable materials were still being removed, sold and recycled by the thousands of informal recyclers who rely on this income for their livelihoods. Waste-picker integration, as outlined in the guidelines approved by the DFFE[34], must gain increased urgency and form a core part of any new waste management systems. Furthermore, any new arrangements must be co-designed with local communities to ensure that services are appropriate and effective for each particular community.

These recommendations require deep structural changes and a dramatic rethinking of our waste management systems. To address inequalities and safeguard our environment, we must think beyond centralised systems hinging on the competence of a failing state. This approach does not look 'beyond' the state but rather towards more inclusive arrangements that involve all actors in the waste value chain, in particular communities and the informal sector. The state should not be released from their constitutional responsibility, but should drive and finance more collaborative approaches that are more inclusive in conception, choice and access. This is not a mandate our municipalities alone can hold, and creating cleaner communities requires meaningful change that must occur at all levels of society.

As recent events in Durban, and the experiences of residents in countless other municipalities across the country, have demonstrated, the consequences of failure are dire, with potentially horrible consequences for our environment and society. Our municipalities may fail, but waste management systems cannot be allowed to fail along with them.

Declarations

We have no competing interests or AI or LLM use to declare. Both authors read and approved the final version of the manuscript.

References

1. National Environmental Management: Waste Act, No. 59 of 2008 [document on the Internet]. c2008 [cited 2024 Jul 10]. Available from: https://www.gov.za/sites/default/files/gcis_document/201409/32000278.pdf
2. Kalina M, Makwetu N, Tilley E. "The rich will always be able to dispose of their waste": A view from the frontlines of municipal failure in Makhanda, South Africa. Environ Dev Sustain. 2024;26:17759–17782. https://doi.org/10.1007/s10668-023-03363-1
3. Schenck CJ, Nell CM, Grobler L, Blaauw PF. Towards engaged solid waste management for cleaner cities and towns in South Africa. Waste Research Development and Innovation Roadmap Research Report. Pretoria: Department of Science and Innovation; 2022.
4. Kanyane M. Exploring challenges of municipal service delivery in South Africa (1994-2013). Afr Public Serv Deliv Perform Rev. 2014;2(1):90–110. https://doi.org/10.4102/apsdpr.v2i1.45
5. Statistics South Africa. Census 2022: Key results. Pretoria: Statistics South Africa; 2022.
6. Erasmus D. Only 38 municipalities receive clean audits – Auditor General SA. Mail & Guardian. 2023 May 31. Available from: https://mg.co.za/article/2023-05-31-only-38-municipalities-receive-clean-audits-auditor-general-sa/
7. Bisseker C. Will debt write-offs stop the rot in delinquent municipalities? Financial Mail. 2023 May 11. Available from: https://www.businesslive.co.za/fm/features/2023-05-11-will-debt-write-offs-stop-the-rot-in-delinquent-municipalities/
8. Navarrete-Hernandez P, Toro F. Urban systems of accumulation: Half a century of Chilean neoliberal urban policies. Antipode. 2019;51(3):899–926. https://doi.org/10.1111/anti.12504
9. Smith A. Sustaining municipal parks in an era of neoliberal austerity: The contested commercialisation of Gunnersbury Park. Environ Plann A. 2021;53(4):704–722. https://doi.org/10.1177/0308518X20951814
10. Schenck CJ, Chitaka TY, Tyrrell H, Couvert A. Disposable diaper usage and disposal practices in Samora Machel township, South Africa. Sustainability. 2023;15(12), Art. #9478. https://doi.org/10.3390/su15129478
11. Viljoen JMM, Schenck CJ, Volschenk L, Blaauw PF, Grobler L. Household waste management practices and challenges in a rural remote town in the Hantam Municipality in the Northern Cape, South Africa. Sustainability. 2021;13(11), Art. #5903. https://doi.org/10.3390/su13115903
12. Kalina M. Waste management in a more unequal world: Centring inequality in our waste and climate change discourse. Local Environ. 2020;25(8):612–618. https://doi.org/10.1080/13549839.2020.1801617
13. Kalina M. As South Africa's cities burn: We can clean-up, but we cannot sweep away inequality. Local Environ. 2021;26(10):1186–1191. https://doi.org/10.1080/13549839.2021.1967900
14. Adeleke O, Akinlabi S, Jen T-C, Dunmade I. Towards sustainability in municipal solid waste management in South Africa: A survey of challenges and prospects. Trans R Soc S Afr. 2021;76(1):53–66. https://doi.org/10.1080/0035919X.2020.1858366
15. Rasmeni ZZ, Madyira DM. A review of the current municipal solid waste management practices in Johannesburg City townships. Procedia Manuf. 2019;35:1025–1031. https://doi.org/10.1016/j.promfg.2019.06.052
16. Haywood LK, Kapwata T, Oelofse S, Breetzke G, Wright CY. Waste disposal practices in low-income settlements of South Africa. Int J Environ Res Public Health. 2021;18(15), Art. #8176. https://doi.org/10.3390/ijerph18158176
17. Mngomezulu SK, Mbanga S, Adeniran AA, Soyez K. Factors influencing solid waste management practices in Joe Slovo Township, Nelson Mandela Bay. J Public Admin. 2020;55(3):400–411.
18. Kalina M, Kwangulero J, Ali F, Abera YG, Tilley E. "Where does it go?": Perceptions and problems of riverine and marine litter amongst South Africa and Malawi's urban poor. PLoS Water. 2022;1(3), e0000013. https://doi.org/10.1371/journal.pwat.0000013
19. Kalina M, Kwangulero J, Ali F, Tilley E. "You need to dispose of them somewhere safe": Covid-19, masks, and the pit latrine in Malawi and South Africa. PLoS ONE. 2022;17(2), e0262741. https://doi.org/10.1371/journal.pone.0262741
20. Slekiene J, Swan N, Kalina M. Absorbent hygiene products disposal behaviour in informal settlements: Identifying determinants and underlying mechanisms in Durban, South Africa. BMC Public Health. 2024;24(1), Art. #912. https://doi.org/10.1186/s12889-024-18396-y
21. Mathekganye J, van Heerden L, Ukwandu D. The nexus between water, neoliberalism and sustainable development in postapartheid South Africa. Afr J Public Aff. 2019;11(3):41–58. https://hdl.handle.net/10520/EJC-19604996c9
22. Yates JS, Harris LM. Hybrid regulatory landscapes: The human right to water, variegated neoliberal water governance, and policy transfer in Cape Town, South Africa, and Accra, Ghana. World Dev. 2018;110:75–87. https://doi.org/10.1016/j.worlddev.2018.05.021
23. Samson M. Whose frontier is it anyway? Reclaimer "integration" and the battle over Johannesburg's waste-based commodity frontier. Capital Nat Social. 2020;31(4):60–75. https://doi.org/10.1080/10455752.2019.1700538
24. Samson M, Kadyamadare G, Ndlovu L, Kalina M. "Wasters, agnostics, enforcers, competitors, and community integrators": Reclaimers, S@S, and the five types of residents in Johannesburg, South Africa. World Dev. 2022;150, Art. #105733. https://doi.org/10.1016/j.worlddev.2021.105733
25. Santos C. Open questions for public water management: Discussions from Uruguay's restatization process. Util Policy. 2021;72, Art. #101273. https://doi.org/10.1016/j.jup.2021.101273
26. Republic of South Africa. Constitution of the Republic of South Africa, No 108. Government Gazette. 1996;378(17678).
27. Baclija Brajnik I, Kronegger L, Prebilic V. Depoliticization of governance in large municipalities in Europe. Urban Aff Rev. 2024;60(2):774–789. https://doi.org/10.1177/10780874231165776
28. Wilson D, Velis CA, Rodic L. Integrated sustainable waste management in developing countries. Waste Resour Manag. 2013;166(2):52–68. https://doi.org/10.1680/warm.12.00005
29. Comins L. The shifting landscape of South Africa's informal settlements. Mail and Gaurdian. 2023 October 30. Available from: https://mg.co.za/news/2023-10-30-the-shifting-landscape-of-south-africas-informal-settlements/
30. Lohri CR, Camenzind EJ, Zurbrügg C. Financial sustainability in municipal solid waste management – costs and revenues in Bahir Dar, Ethiopia. Waste Manag. 2014;34(2):542–552. https://doi.org/10.1016/j.wasman.2013.10.014
31. Mokgabodi B. Funding municipal waste management services [document on the Internet]. c2013 [cited 2024 Jul 10]. Available from: https://www.dffe.gov.za/sites/default/files/docs/national_treasury_funding_waste_services.pdf
32. Coffey M, Coad A. Collection of municipal solid waste in developing countries [document on the Internet]. c2010 [cited 2024 Jul 10]. Available from: https://unhabitat.org/sites/default/files/2021/02/2010_collection-msw-developing-countries_un-habitat.pdf
33. Godfrey L, Ahmed MT, Gebremedhin KG, Katima JK, Oelofse S, Osibanjo O, et al. Solid waste management in Africa: Governance failure or development opportunity? In: Edomah N, editor. Regional development in Africa. London: IntechOpen; 2020. https://doi.org/10.5772/intechopen.86974
34. Department of Environment, Forestry and Fisheries (DEFF) and Department of Science and Innovation (DSI). Waste picker integration guideline for South Africa: Building the recycling economy and improving livelihoods through integration of the informal sector. Pretoria: DEFF and DSI; 2020.

AUTHOR:
Naomi Hlongwane[1]

AFFILIATION:
[1]Centre for Disability and Rehabilitation Studies, Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

CORRESPONDENCE TO:
Naomi Hlongwane

EMAIL:
nhlongwane@setshaba.org.za

HOW TO CITE:
Hlongwane N. Navigating the complexities of long-term care service delivery for older people in South Africa. S Afr J Sci. 2024;120(11/12), Art. #18721. https://doi.org/10.17159/sajs.2024/18721

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
long-term care, older people, quality of life, service delivery, long-term care policy

PUBLISHED:
30 August 2024

Navigating the complexities of long-term care service delivery for older people in South Africa

Significance:

This Commentary addresses the fundamental difficulties in providing equitable and sustainable long-term care services to older people in South Africa. It emphasises the long-term consequences of apartheid and structural racial inequities, as well as the necessity for a comprehensive regulatory framework and integrated policies. The critical need for workforce development, proper funding, and the integration of health and social care services are also highlighted. The goal is to increase service delivery efficacy, improve the quality of life for older people, and promote sustainability in the long-term care sector by suggesting actionable solutions.

Introduction

The World Health Organization (WHO) recommends that all countries have a long-term care (LTC) system, defining LTC broadly as: "Activities undertaken by others to ensure that people with or at risk of a significant ongoing loss of intrinsic capacity can maintain a level of functional ability consistent with their basic rights, fundamental freedoms and human dignity."[1]

The WHO further stipulates that LTC must be affordable and easily accessible, with special attention paid to ensuring that the marginalised have access to services. With the increasing population, there is a greater need to focus on LTC for older people as the demand is likely to increase.[2,3] The provision of LTC services for older people (an 'older person' means a person who is 60 years of age or older according to the *Older Persons Act*);in South Africa presents a multifaceted challenge that requires immediate attention.[4,5] The care and experiences of older people in LTC facilities such as nursing homes and assisted living communities reflect historical legacies, systemic deficiencies, and socio-economic disparities.[6-8] The aim of this Commentary is to provide a critical examination of the intricate dimensions of LTC service delivery in South Africa, highlighting disparities, regulatory frameworks, workforce capacity, and funding challenges. Based on scientific insights and empirical evidence, I propose actionable solutions to enhance service delivery effectiveness and promote sustainability.

Unravelling the complexities

Racial disparities and historical legacies

The legacy of apartheid policies in South Africa, a historical system of institutionalised racial segregation and discrimination until the early 1990s, continues to influence contemporary society. For many years before and during the apartheid era, there was a clear disparity in service provision, healthcare access and socio-economic status among different ethnic groups in South Africa.[9] Specifically, residential care facilities were accessible only to older white individuals during that period. However, with the end of apartheid in 1994 and the implementation of the Transformation Act, access to these facilities became available to all racial groups.[10] Despite this transformed access for all racial groups, residential care in African contexts has often been viewed as unconventional due to its perceived contradiction with traditional African values, which centre on family-based care.[11,12] This view is supported by the existing research, which indicates a prevalent belief in African communities that old age homes primarily accommodate childless older people or those whose families neglect them and leave them with no alternative.[13] Studies conducted in the last decade, though, show that there has been a change in attitudes towards the family-centred model of care for older people in sub-Saharan African countries, as more people are making use of the various LTC facilities. This change is being driven partly by urbanisation and industrialisation.[14,15]

The impact of HIV/AIDS and dementia on LTC

In South Africa, despite the change in care models, the enduring legacy of apartheid compounded by the HIV/AIDS epidemic continues to cast a long shadow over LTC provision in South Africa. The impact of HIV in LTC is multifaceted, especially as the population of older people living with HIV has increased due to improved antiretroviral therapy (ART) and life expectancy.[16,17] Studies show that while ART has extended life expectancy, it also leads to various chronic conditions such as cardiovascular disease, low bone mineral density, and increased risk of falls.[18,19] Additionally, ageing with HIV presents unique challenges in LTC settings, with higher rates of dementia, antipsychotic use, and potential adverse effects from polypharmacy, including drug interactions and neurocognitive side effects.[20] The adoption of universal HIV treatment guidelines has shown lower retention rates after ART initiation, emphasising the need for continuous monitoring of long-term care outcomes to address gaps in care quality and patient attrition for older people.[21]

Notwithstanding the impact of HIV alone, older people living with dementia in LTC facilities in South Africa face various challenges, including socio-economic deprivation, lack of appropriate care and support services, and barriers to diagnosis and treatment.[22] Additionally, the lack of advanced technology and appropriate medication in LTC facilities poses challenges in providing adequate care for patients with dementia.[23] Furthermore, structural factors create barriers to diagnosis, support, and care for people living with dementia in South Africa, highlighting the urgent need for intersectoral policy responses to strengthen health and social care systems for patients with dementia and their families.[24]

As highlighted by Lloyd-Sherlock's seminal work, racial disparities in service delivery are deeply entrenched and contribute to the gaps in service availability and uneven service quality.[25] Beyond these disparities, there is a significant lack of recognition for the necessity of an overarching LTC policy and system. This gap stems from a poor understanding and inadequate attention to the needs of older people, particularly those living with dementia.

Regulatory frameworks and coordination challenges

The legal and policy framework governing LTC in South Africa is anchored by the *Older Persons Act of 2006* (OPA), complemented by broader legal and policy instruments including The United Nations Decade of Healthy Ageing (2021–2030) that aligns healthy ageing with the United Nations' Sustainable Development Goals (SDGs).[26] Despite the policy framework mapping out inclusivity and reform, studies conducted in South Africa inform us that only a minority of older people receive publicly funded LTC.[25,27] Adding to the reflected historical racial divides, most facilities are also managed by non-governmental and faith-based organisations, and the standard of care in these facilities is highly variable.[28] Reflecting on data from the 2010 audit of residential facilities for older people in South Africa (conducted by Umhlaba Development Services), reveals that residential services for older people remain one of the most immutable areas in the provision of social welfare services. Although services for older people were simplified and de-racialised nearly 20 years ago, it was found at the time of the audit, that black and white residents of 10 facilities were physically separated and residents did not appear to receive the same quality and service standards.[29] The coordination and implementation of these regulatory frameworks mentioned above as well as policies primarily fall within the purview of the Department of Social Development (DSD) as indicated in the audit, and the Department of Health. The former administers old-age pensions and finances and oversees residential, community and home-based care, while the latter addresses older people's healthcare needs. Overall, coordination of LTC across these departments is lacking and clinical-level integration of health and social care is limited. This fragmentation hinders the development of comprehensive and cohesive strategies to address the multifaceted needs of older people and the delivery of much-needed services.

Workforce capacity and training deficits

Amidst this burgeoning ageing population in South Africa, concerns have been raised regarding the healthcare system's capacity to meet escalating demands, particularly within LTC settings.[30,31] Compounding the problem is the shortage of trained and experienced nurses and care workers, which hampers the sector's ability to provide adequate care. Evidence suggests that the current workforce lacks adequate training to address the evolving needs of the ageing population effectively.[32-34] A large number of learning goals for older people are included in medical and nursing training programmes; however, it is with limited coverage and lack of discrete assessment in this area.[32] This further perpetuates the notion that older people are not prioritised in the South African agenda, thus hindering their quality of life, social inclusion, and LTC service delivery.[33,35] A South African national survey of 405 supervised long-term care facilities conducted in 2010 found that only a quarter of the staff knew about official norms and standards pertaining to the quality of care and life of older people.[36] This finding was later reflected in a report written by Makgoba on the unlawful deaths and catastrophic transfer of residents at the Life Esidimeni Centre.[37] Professional training for care workers is often inadequate, inconsistent, and unaccredited, as reflected in Makgoba's work, and the survey conducted by DSD furthermore revealed that care workers' skills certificates are not acknowledged for career progression, leading to job dissatisfaction and viewing care work as a "career-less" job. The African Union advocates public policies that ensure care workers have the skills and knowledge required for their roles.[38] In addition to advocacy efforts, the South African government enacted the OPA and its regulations, and set policies through the Health and Safety Sector Education and Training Authority to provide training and support for care workers. However, a career pathway and accreditation for care workers within the LTC sector is still lacking.[39] The Policy on Social Service Practitioners intended to regulate care work and create an occupational framework for care workers; however, this too fell short of implementation.[40,41] The lack of implementation has left care workers grappling with challenges due to insufficient government recognition and policy support.[40,42,43]

Dwindling funds impact LTC service delivery

In the backdrop of these challenges, during the parliamentary monitoring committee meeting on 27 March 2024, the committee reviewed the legacy report on the DSD's quarterly performance. The committee highlighted that the older persons programme is experiencing a financial crisis, exacerbated by a ZAR16 million budget cut in the last financial year. This funding shortfall has led to the closure of several older persons' service centres. Over and above the budget cuts, the government social grants for older people and subsidies for LTC facilities are not keeping pace with rising costs, forcing many facilities to close their doors. Financial struggles are widespread among organisations that provide social care services across South Africa[44,45], and the pandemic has exacerbated these challenges, curtailing fundraising activities and increasing financial pressure.[46] Many non-profit organisations and associations find their funds inadequate to cover the escalating costs of services, leading to growing deficits and further closures. Two reports from the Family Caregiving of Older Persons in South Africa Programme highlight that most funding is allocated to the Older Persons Grant, which benefits 75% of the older people population. However, only 2% of the funding supports community-based and residential care, reaching a small fraction of those in need.[47,48] There are also vast disparities in funding allocation. There is a need to review the funding model for LTC for older people in order to improve the sustainability and quality of older people care in South Africa.[49]

Enhancing LTC service delivery

To enhance the quality of LTC, policies should prioritise increased public spending and improved regulations, including establishing quality assessment and continuous improvement monitoring systems. While governments in many countries are taking a more proactive approach, LTC within South Africa still lags behind acute health care in terms of measurement and quality improvement strategies.[50] Addressing this gap requires greater investment, with standards focusing on improving outcomes rather than just infrastructure. A critical issue the South African government needs to address is the gaps in staffing and training. Current staffing levels in LTC are insufficient to support sustainable service delivery in the future. Furthermore, the significant shortage of trained nurses and care workers compromises the ability to deliver sufficient care. Existing workforce training fails to adequately address the evolving needs of older adults, perpetuating a cycle in which older people care is marginalised within national priorities. Organisational shortcomings, such as a lack of adherence to quality norms and insufficient policy frameworks, further hinder the provision of effective LTC services. Addressing these shortages and gaps in implementation now is crucial to prevent further deterioration of older people's quality of life. Moreover, to promote equitable access to care and improve the quality of life for older persons, inclusive policies, tailored interventions, and community-based activities are essential. Investing in community-based services can improve access for older people in underserved areas, fostering an inclusive and equitable system, as there is a growing move towards 'ageing in place' as an alternative model of care to institutional living.[51] Investing in LTC requires adequate funding models and this includes moving away from relying solely on government funds and instead integrating private contributions and involvement. Establishing a robust policy on public-private partnerships is essential for strategically financing LTC services for older people. This approach could adequately prepare for an ageing population by engaging a wide range of stakeholders and utilising resources from the private sector. A scoping review by Aghdash and colleagues found that public-private partnerships enhance older people's access to healthcare services, improve care effectiveness, and reduce costs for both health systems and older people.[52] Previous implementation of public-private partnerships shows valuable insights into best practices for future initiatives, such as enhancing LTC service delivery and addressing financial challenges. Applying these principles

could similarly combine private sector investments with government funding to enhance access to the quality of older people's care.

Conclusion

The delivery of LTC services to older adults in South Africa is a complex endeavour that necessitates a varied and collaborative approach. South Africa can improve service delivery effectiveness and promote sustainability in the LTC sector by implementing evidence-based policies, investing in infrastructure and workforce development, supporting family caregivers and community-based care, and encouraging collaboration among various stakeholders.

Recommendations

Call to action: Formulate a comprehensive long-term care policy

There is an urgent need for a comprehensive LTC policy to ensure the needs of older people are recognised and systematically addressed. The OPA, although including components of residential care, does not address the needs of older people within the LTC system, and thus the recommendation for a call for a dedicated LTC policy. This policy should integrate LTC into national health strategies, social protection systems, and development plans, providing a robust regulatory framework to guide these initiatives effectively. It should support family and community-based care by establishing a system that prioritises home-based care and strengthens community services, including respite care, day care centres, and stimulation programmes. While residential care should be available, it should not be the primary focus. The policy must also address the needs of older people requiring LTC due to severe mental illness, intellectual disabilities, or mobility impairments. By addressing the needs of older people, the policy should be inclusive of the required training by prioritising the training and professional development of care workers, acknowledging their critical role in LTC. A well-trained, supported workforce is essential for delivering high-quality care and for promoting compliance with standards. The policy proposed should clearly stipulate the implementation process of upholding the rigorous quality standards for LTC services across all care settings, including family, community-based, and residential facilities. This ensures a consistent level of care and supports the well-being of older people receiving LTC. Beyond this proposed LTC policy, data on the functional health and socio-economic status of older adults need to be gathered systematically. This information is vital for effective planning, resource allocation, and the development of responsive, relevant LTC services that meet the actual needs of the population. Filling these gaps is crucial to ensuring the quality of life and dignity of older people and their caregivers. By also investing in research and developing culturally relevant, context-specific care models, we can better support older people and caregivers, ensuring their basic rights and fundamental freedoms are upheld.

It is imperative for government, civil society, and academia to collaborate in leading the development and implementation of LTC policies. This collective effort will support an often-overlooked sector of society, ensuring a dignified and well-supported life for older people in South Africa.

Declarations

I have no competing interests to declare. AI was not used in the writing or editing of this manuscript.

References

1. World Health Organization (WHO). World report on ageing and health. Geneva: WHO; 2015. Available from: https://iris.who.int/bitstream/handle/10665/186463/?sequence=1
2. Lloyd-Sherlock P, Pot AM, Sasat S, Morales-Martinez F. Volunteer provision of long-term care for older people in Thailand and Costa Rica. Bull World Health Organ. 2017;95(11), Art. #774. https://doi.org/10.2471/BLT.16.187526
3. Kim H, Kwon S. A decade of public long-term care insurance in South Korea: Policy lessons for aging countries. Health Policy. 2021;125(1):22–26. https://doi.org/10.1016/j.healthpol.2020.11.003
4. Rankin SM. Older adults' experiences of formal community-based care services in Sebokeng: Implications for long-term care management [thesis]. Potchefstroom: North West University; 2019. https://repository.nwu.ac.za/bitstream/handle/10394/33786/Rankin_M.pdf?sequence=1
5. Lazarus JV, Safreed-Harmon K, Kamarulzaman A, Anderson J, Leite RB, Behrens G, et al. Consensus statement on the role of health systems in advancing the long-term well-being of people living with HIV. Nat Commun. 2021;12(1), Art. #4450. https://doi.org/10.1038/s41467-021-24673-w
6. Mauldin RL, Sledge SL, Kinney EK, Herrera S, Lee K. Addressing systemic factors related to racial and ethnic disparities among older adults in long-term care facilities. In: Guerrero E, editor. Effective elimination of structural racism. London: IntechOpen; 2022. https://doi.org/10.5772/intechopen.99926
7. Schatz E, Madhavan S, Collinson M, Gómez-Olivé FX, Ralston M. Dependent or productive? A new approach to understanding the social positioning of older South Africans through living arrangements. Res Aging. 2015;37(6):581–605. https://doi.org/10.1177/0164027514545976
8. Aboderin I, Kano M, Vincent HA. Toward "age-friendly slums"? Health challenges of older slum dwellers in Nairobi and the applicability of the age-friendly city approach. Int J Environ Res Public Health. 2017;14(10), Art. #1259. https://doi.org/10.3390/ijerph14101259
9. McLaren ZM, Ardington C, Leibbrandt M. Distance decay and persistent health care disparities in South Africa. BMC Health Serv Res. 2014;14:1–9. https://doi.org/10.1186/s12913-014-0541-1
10. Westaway MS, Olorunju SA, Rai L-CJ. Which personal quality of life domains affect the happiness of older South Africans? Qual Life Res. 2007;16:1425–1438. https://doi.org/10.1007/s11136-007-9245-x
11. Aboderin IA, Beard JR. Older people's health in sub-Saharan Africa. Lancet. 2015;385(9968):e9–e11. https://doi.org/10.1016/S0140-6736(14)61602-0
12. Pype K. Caring for people 'without' value: Movement, reciprocity and respect in Kinshasa's retirement homes. In: Hoffman J, Pype K, editor. Ageing in sub-Saharan Africa: Spaces and practices of care. Bristol: Policy Press; 2016. p. 43–70. https://doi.org/10.1332/policypress/9781447325253.003.0003
13. Ncube N. Pathways to institutional care for elderly indigenous Africans: Navigating contours of alternatives. Afr J Soc Work. 2017;7(1):44–51. Available from: https://www.ajol.info/index.php/ajsw/article/view/160528
14. Budlender D, Lund F. South Africa: A legacy of family disruption. Dev Change. 2011;42(4):925–946. https://doi.org/10.1111/j.1467-7660.2011.01715.x
15. Kagitcibasi C, Ataca B. Value of children, family change, and implications for the care of the elderly. Cross-Cult Res. 2015;49(4):374–392. https://doi.org/10.1177/1069397115598139
16. Ahmed MH, Ahmed F, Abu-Median A-B, Panourgia M, Owles H, Ochieng B, et al. HIV and an ageing population-what are the medical, psychosocial, and palliative care challenges in healthcare provisions. Microorganisms. 2023;11(10):24–26. https://doi.org/10.3390/microorganisms11102426
17. Jaqua E, Labib W, Danji K. HIV-associated conditions in older adults. Cureus. 2022;14(12), e32661. https://doi.org/10.7759/cureus.32661
18. Althoff KN, Smit M, Reiss P, Justice AC. HIV and ageing: Improving quantity and quality of life. Curr Opin HIV AIDS. 2016;11(5):527–536. https://doi.org/10.1097/COH.0000000000000305
19. Erlandson KM, Karris MY. HIV and aging: Reconsidering the approach to management of comorbidities. Infect Dis Clin North Am. 2019;33(3):769–786. https://doi.org/10.1016/j.idc.2019.04.005
20. Justice AC, Akgün KM. What does aging with HIV mean for nursing homes? J Am Geriatr Soc. 2019;67(7), Art. #1327. https://doi.org/10.1111/jgs.15950
21. Brazier E, Tymejczyk O, Wools-Kaloustian K, Jiamsakul A, Torres MTL, Lee JS, et al. Long-term HIV care outcomes under universal HIV treatment guidelines: A retrospective cohort study in 25 countries. PLoS Med. 2024;21(3), e1004367. https://doi.org/10.1371/journal.pmed.1004367
22. Van Wyk A, Manthorpe J, Clark C. The behaviours that dementia care home staff in South Africa find challenging: An exploratory study. Dementia. 2017;16(7):865–877. https://doi.org/10.1177/1471301215622092
23. Kantawala B, Zahra N, Oseili T, Jawad A, Ouardouz S, Nazir A, et al. Dementia challenges in Africa: Journeying through forgetfulness. Health Sci Rep. 2023;6(9), e1561. https://doi.org/10.1002/hsr2.1561

24. Jacobs R, Schneider M, Farina N, du Toit P, Docrat S, Comas-Herrera A, et al. Dementia in South Africa: A situational analysis. Dementia. 2024;23(3):452–475. https://doi.org/10.1177/14713012231183358

25. Lloyd-Sherlock P. Long-term care for older people in South Africa: The enduring legacies of apartheid and HIV/AIDS. J Soc Policy. 2019;48(1):147–167. https://doi.org/10.1017/S0047279418000326

26. Shevelkova V, Mattocks C, Lafortune L. Efforts to address the Sustainable Development Goals in older populations: A scoping review. BMC Public Health. 2023;23(1):456. https://doi.org/10.1186/s12889-023-15308-4

27. Kalideen L, Van Wyk JM, Govender P. Demographic and clinical profiles of residents in long-term care facilities in South Africa: A cross-sectional survey. Afr J Prim Health Care Fam Med. 2022;14(1), Art. #3131. https://doi.org/10.4102/phcfm.v14i1.3131

28. Marx FJ. Exploring alternative residential care facilities for the intermediate elder: Towards a retirement facility in Warwick Junction [dissertation]. Durban: University of KwaZulu-Natal; 2016. http://hdl.handle.net/10413/15260

29. South African Department of Social Development. Audit of residential facilities [document on the Internet]. c2010 [cited 2024 May 20]. Available from: https://social.un.org/ageingworkinggroup/documents/FINAL%20REPORT%20DSD%20Audit%20of%20Residential%20Facilities%20April2010.pdf

30. Goodrick WF. Policy implications and challenges of population ageing in South Africa [dissertation]. Bloemfontein: University of the Free State; 2013. https://scholar.ufs.ac.za/server/api/core/bitstreams/38d9b5e1-3c94-433e-a933-731e0027eac9/content

31. Dowling E. The care crisis: What caused it and how can we end it? London: Verso; 2022.

32. Naidoo K, Van Wyk J. What the elderly experience and expect from primary care services in KwaZulu-Natal, South Africa. Afr J Prim Health Care Fam Med. 2019;11(1), Art. #2100. https://doi.org/10.4102/phcfm.v11i1.2100

33. Kelly G, Mrengqwa L, Geffen L. "They don't care about us": Older people's experiences of primary healthcare in Cape Town, South Africa. BMC Geriatr. 2019;19:1–14. https://doi.org/10.1186/s12877-019-1116-0

34. Mapira L, Kelly G, Geffen LN. A qualitative examination of policy and structural factors driving care workers' adverse experiences in long-term residential care facilities for older adults in Cape Town. BMC Geriatr. 2019;19:1–8. https://doi.org/10.1186/s12877-019-1105-3

35. Kang'ethe S. Violation of human rights of older persons in South Africa. The case of Lavela old age centre, Ntselamanzi, Eastern Cape Province, South Africa. Soc Work. 2018;54(3):283–295. https://doi.org/10.15270/54-3-649

36. Soouth African Department of Social Development. Audit of residential facilities [document on the Internet]. c2010 [cited 2024 May 20]. Available from: https://social.un.org/ageing-working-group/documents/FINAL%20REPORT%20DSD%20Audit%20of%20Residential%20Facilities%20April2010.pdf

37. Makgoba M. The report into the circumstances surrounding the deaths of mentally ill patients: Gauteng Province. Pretoria: Office of the Health Ombud; 2017. https://www.sahrc.org.za/home/21/files/Esidimeni%20full%20report.pdf

38. African Union. AU policy framework and plan of action on ageing. Cooperation between Help Age International Africa Regional Development Centre and the African Union. Nairobi: Kenya Help Age International; 2002. http://www.un.org/esa/socdev/ageing/documents/implementation/AUFrameworkBook.pdf

39. Patel L. The gendered character of social care in the non-profit sector in South Africa. Johannesburg: Centre for Social Development in Africa, UNRISD and Centre for Social Development Research in Africa, University of Johannesburg; 2009. https://ujcontent.uj.ac.za/esploro/outputs/9911597407691

40. Gray M, Lombard A. Progress of the social service professions in South Africa's developmental social welfare system: Social work, and child and youth care work. Int J Soc Welf. 2023;32(4):429–441. https://doi.org/10.1111/ijsw.12562 Accessed 2024 May 20.

41. South African Department of Social Development (DSD). National department of social development, South Africa. Policy on social service practitioners. Pretoria: DSD; 2017 [cited 2024 May 20]. https://www.westerncape.gov.za/assets/departments/social-development/040.pdf

42. Erasmus C, Ramuhaheli R. Challenges experienced while providing home based care: A community organisation's experience. J Community Health Sci. 2012;7(1):21–27.

43. The Health and Welfare Sector Education and Training Authority, South Africa. Sector skills plan update for the health and social development sector in South Africa. HWSETA SSP update 2014–2015 for the period 2015–2020. Pretoria: Department of Higher Education and Training; 2015 [cited 2024 May 20]. https://www.wrseta.org.za/sites/default/files/2020-01/SSP_2014_15.pdf

44. Sibisi N, Makka A. Financial challenges experienced by South African non-profit organisations regarding CSR implementation. Soc Responsib J. 2021;18(6):1089–1105. https://doi.org/10.1108/SRJ-12-2019-0397

45. Ijon R, Azman A, Singh J, Singh P. Funding sustainability of welfare oriented non-governmental organisations: A brief analysis. Int Online J Lang Commun Humanit. 2021;4(1):92–106.

46. Hlongwane N, Ned L, McKinney E, McKinney V, Swartz L. Experiences of organisations of (or that serve) persons with disabilities during the COVID-19 pandemic and national lockdown period in South Africa. Int J Environ Res Public Health. 2022;19(19), Art. #12641. https://doi.org/10.3390/ijerph191912641

47. Moore E, Kelly G. Funding elder care in South Africa report. Cape Town: University of Cape Town; 2023. https://doi.org/10.25375/uct.24146772.v3

48. Moore E, Kelly G. Policy brief 2: Addressing the gaps in long-term care for older people in southern Africa. Cape Town: University of Cape Town; 2023 [cited 2024 May 20]. https://familycaregiving.org.za/wp-content/uploads/2023/04/Long-Term-Care-for-Older-Policy-Brief.pdf

49. World Health Organization (WHO). Financing long-term care programmes in health systems, with a situation assessment in selected high-, middle-and low-income countries. Geneva: WHO; 2007. https://iris.who.int/bitstream/handle/10665/85676/HSS_HSF_DP.07.6_eng.pdf

50. World Health Organization (WHO). Rebuilding for sustainability and resilience: Strengthening the integrated delivery of long-term care in the European Region. Geneva: WHO Regional Office for Europe; 2022. https://iris.who.int/bitstream/handle/10665/353912/WHO-EURO-2022-5330-45095-64318-eng.pdf?sequence=1

51. Barati F, Khoddam H, Modanloo M. Concept development of "Aging in place": Application of hybrid model. J Family Med Prim Care. 2022;11(11):6646–6653. https://doi.org/10.4103/jfmpc.jfmpc_2150_21

52. Aghdash SA, Sharifzade H, Bande-Ehahi K, Bolbanabad AM, Gharaee H. Public-private partnership experiences and achievements in providing primary health care to the elderly: A scoping review. Health Technol Assess Action. 2021;5(3):2–14. https://doi.org/10.18502/htaa.v5i3.9346

AUTHOR:
Alecia Samuels[1]

AFFILIATION:
[1]Centre for Augmentative and Alternative Communication, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Alecia Samuels

EMAIL:
alecia.samuels@up.ac.za

HOW TO CITE:
Samuels A. Stuck in old ways: Towards transdisciplinary training for healthcare professionals in early intervention. S Afr J Sci. 2024;120(11/12), Art. #18750. https://doi.org/10.17159/sajs.2024/18750

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
early intervention, transdisciplinarity, interprofessional education, rehabilitation, National Health Insurance

Stuck in old ways: Towards transdisciplinary training for healthcare professionals in early intervention

Significance:

South Africa has a serious shortage of healthcare professionals trained to provide early intervention services for young children with disabilities, especially in the public healthcare sector, accessed by the majority of the South African population. A transdisciplinary approach which requires healthcare professionals to work across and outside their disciplinary boundaries is recommended, but few healthcare professionals in South Africa are trained in this manner. A need exists, therefore, to transform the training of healthcare professionals to ensure that more children with disabilities can receive early intervention services.

Introduction

Like many other low- and middle-income countries on the African continent, South Africa has a significant shortage of healthcare professionals.[1] Across the 47 countries of the WHO African region, there are only approximately 3.6 million healthcare workers, the majority of whom are nurses and medical doctors.[2] A much smaller percentage of this workforce is made up of rehabilitation professionals such as occupational therapists (OTs), physiotherapists (PTs), speech-language therapists (SLTs), and audiologists (AUDs), amongst others.[2]

In South Africa, the result of this workforce shortage means that access to rehabilitation services for those with short- and long-term disabilities is desperately lacking, especially in the public healthcare sector[1] and even more so in rural areas[3]. For example, in 2022, there were 6063 OTs, 8571 PTs, and 4072 SLTs and AUDs registered with the Health Professions Council of South Africa.[4] Within South Africa's two-tiered healthcare system, of these, 1101 (18%) OTs, 1224 (14%) PTs, and 617 (15%) SLTs and AUDs were employed in the public healthcare sector[4], with the rest presumably operating in private health care or education, or in other countries, or possibly unemployed or awaiting placement. The public sector workforce equates to 2.8 OTs, 3.2 PTs, and 1.7 SLTs per 100 000 patients in the public healthcare sector[5] which services 84% of the South African population[1]. Only 40% of the public sector rehabilitation therapists are estimated to work in rural settings, with inexperienced community service therapists making up a third of this rural workforce.[1] The private healthcare setting, which is estimated to service 16% of the South African population, employs the majority of the health and rehabilitation workforce in the country.[1] Such stark inequality between public and private health care is one of the main drivers for South Africa's National Health Insurance (NHI) Bill[6] which has recently been signed into law.

Children with disabilities typically require a range of therapeutic and rehabilitative interventions from a range of rehabilitation professionals due to the nature of their disability.[3] For the country's estimated 1.15 million young (0–5 years) children with sensory, cognitive and physical disabilities[6], personnel shortages, especially in rural areas, effectively mean that many will miss out on important therapeutic services[3] that could enhance their development or prevent further delays[7].

Over and above workforce constraints, the multidisciplinary and siloed nature in which many rehabilitation professionals operate to deliver early intervention services in South Africa[7,8] makes inefficient and ineffective use of already scarce personnel resources[3] and undermines the quality of services to young children with disabilities. Multidisciplinary rehabilitation teams consist of various disciplinary members that individually assess a child, carry out interventions, and write separate reports and goals within their own disciplinary boundaries.[9] For caregivers of young children with disabilities who are fortunate to gain access to these scarce early intervention services in the public sector, a multidisciplinary approach often requires them to consult separately and at different times, or possibly even different days, with a range of professionals like OTs, PTs and SLTs.

This approach is inefficient, especially for poor families for whom travel to hospitals for different appointments can cut into already constrained financial resources.[7] Furthermore, when rehabilitation professionals work independently of each other in this way, service delivery becomes ineffective as professionals tend to share minimal information about the child between themselves. This may lead to diverse and even contradictory intervention goals which can be confusing for families of children with disabilities.[9]

The multidisciplinary approach also operates on the incorrect assumption that professional services directed to children with disabilities provided once a week or a month within clinical settings are the main pathways through which children learn. This minimises the role that primary caregivers of children with disabilities play in supporting their children's development. Ecological theories and evidence-based research have shown that primary caregivers are the most important influence on their children's development, as learning occurs in the context of children's participation in everyday routines and activities in their natural environments.[10] This does not negate the need for professional rehabilitation services but merely emphasises that important primary caregivers who spend most of their time with children have many more opportunities to influence children's learning and development outside the therapeutic setting. Mahoney and Wiggers[10] estimate this influence to be in the region of 12 times more than that of rehabilitation therapists.

In contrast, the more efficient transdisciplinary approach to service delivery has been put forward as an alternative collaborative model of service delivery[3] in early childhood intervention (ECI)[9], especially within the South African context for children with disabilities[7]. The transdisciplinary approach encourages more dynamic collaboration between rehabilitation disciplines, requiring professionals to think and work outside their disciplinary boundaries to establish collective goals together with families of children with disabilities. Within the transdisciplinary approach, families of children with disabilities work with only one professional who acts as the primary service provider or case manager and becomes the family's link with the rest of the team of rehabilitation professionals.[9]

While there have been previous calls for a move towards a transdisciplinary approach in the rehabilitation professions in South Africa[3,7], few of these calls have shown how this can be done in practice. In the context of rehabilitation personnel shortages within the public healthcare sector, this Commentary therefore focuses specifically on how service delivery for children with disabilities can be transformed by incorporating the transdisciplinary approach into the education and training of rehabilitation professionals within undergraduate, postgraduate and continuing education programmes. This would be one way to ensure more effective ECI services in South Africa for children with disabilities, in line with the NHI's commitment to addressing equitable health care for all South Africans.

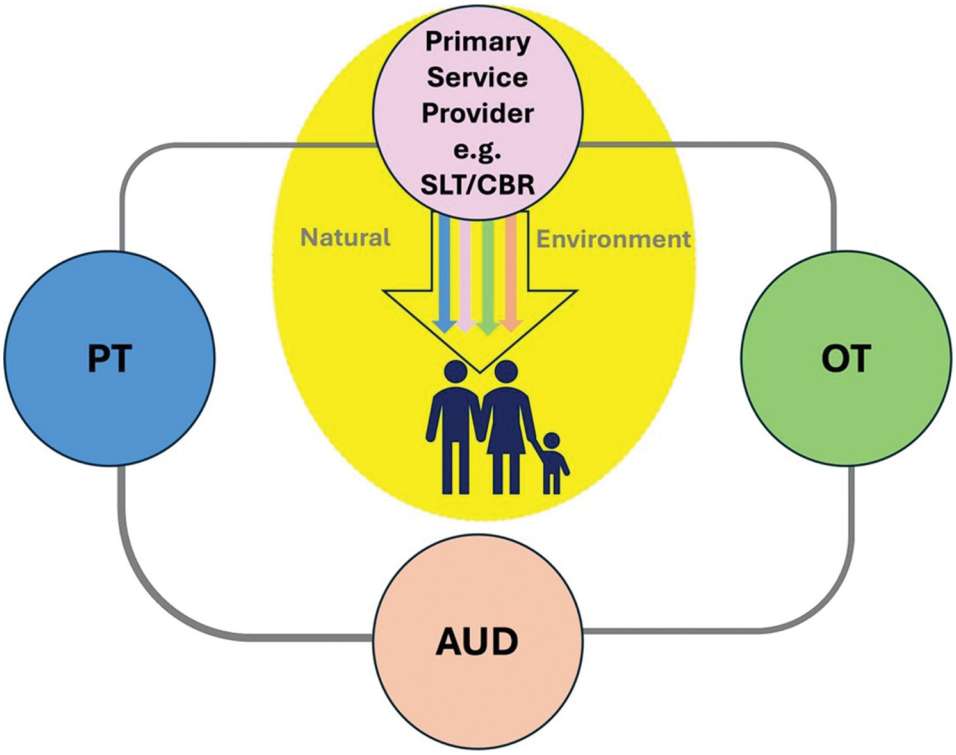
Understanding the transdisciplinary service delivery model for ECI

The transdisciplinary model for service delivery is the recommended practice in the field of ECI for children with disabilities as it decreases the fragmentation of service delivery while enhancing communication between team members and service coordination.[9] This model involves a team of rehabilitation professionals, mainly OTs, SLTs and PTs, but can also involve additional members of the healthcare team, e.g. community-based rehabilitation workers.[3] In less-resourced rural settings, for example, community-based rehabilitation workers (rehabilitation personnel without professional degrees) may represent a cost-effective resource for transdisciplinary teams[3] as they have skills in working at a community and household level under the supervision of professionally trained rehabilitation practitioners[11].

Transdisciplinary team members thus commit to teach, learn and work together across disciplinary boundaries to implement a unified intervention plan.[9] In contrast to the multidisciplinary model, families are valued members of the team in a transdisciplinary model and are involved from the very outset in assisting with the development of an intervention plan based on the family's needs and resources.[10]

In terms of implementation, one member is designated by the team as the primary service provider for a particular case and acts as the link between the rehabilitation team and the family, as depicted in Figure 1. This then enables the rest of the team to become primary service providers for other cases. However, the full team of professionals remains involved with the case as the primary service provider reports back to the team frequently and is trained by those from the other disciplines to implement the agreed-upon intervention plan through the process of role release (described later in Table 1).

The rest of the team may remain involved in other ways, including periodic joint visits for consultation or monitoring purposes. They are also responsible for providing ongoing oversight of their disciplinary aspect for the particular child, and for coaching the primary service provider in discipline-specific strategies that are relevant to the child and family in realising the team's intervention goals. Each team member, therefore, still maintains ethical responsibility for the case with the primary service provider mainly acting as the 'hands' through which the team's intervention goals are realised. This way of working does not fall outside the scope of practice of the rehabilitation professions in South Africa. The regulations defining the scope of the profession of speech-language therapy[12], for example, promote SLT collaboration with other professionals, serving as case managers and service delivery coordinators, as well as providing in-service training to families, caregivers and other professionals. This is also congruent with international speech therapy bodies such as the American Speech-Language-Hearing



OT, occupational therapist; AUD, audiologist; PT, physiotherapist; SLT, speech-language therapist; CBR, community-based rehabilitation worker

Figure 1: The transdisciplinary team approach.

Association (ASHA), in which transdisciplinary collaboration within the context of Interprofessional Education and Practice is promoted.[13]

It should be understood, however, that complex interventions by a specific discipline that pose a risk to clients (e.g. SLTs offering dysphagia intervention) should not be delegated to other team members who are not SLTs.[9]

Moving from multidisciplinary to transdisciplinary ECI

The multidisciplinary approach to ECI service delivery for children with disabilities in South Africa is still firmly entrenched due to the traditional, unidisciplinary medical model of undergraduate training in the rehabilitation disciplines.[9] Curriculum content and clinical practicum experiences at an undergraduate level tend therefore to be focused on a specific discipline. Unidisciplinary education within the rehabilitation sciences facilitates power hierarchies and competition while actively discouraging cooperation between disciplines.[14] This then ultimately filters down to clinical practice in South Africa where rehabilitation professionals still struggle to work together.[14]

Currently, the only opportunity for rehabilitation professionals in South Africa to gain any formal training on the transdisciplinary approach is at a postgraduate level through an Interprofessional Education (IPE) master's programme in ECI at the University of Pretoria. The programme started in 2001[15] and has trained approximately 350 graduated rehabilitation professionals in the transdisciplinary approach. However, this number should be acknowledged as a mere fraction of those registered to practice. Alongside gaining knowledge and skills in transdisciplinary teamwork, an emphasis is also placed on incorporating the biopsychosocial framework of the World Health Organization's (WHO) International Classification of Disability, Functioning and Health, to gain an understanding of the child's ability to participate in everyday routines and activities. Participation within home and education routines as well as in activities in community settings with their caregivers, is the context within which children's learning and development takes place.[10]

The International Classification of Disability, Functioning and Health thus becomes the vehicle through which those trained in the transdisciplinary approach can develop shared goals. This is a requirement for collaborative goal setting in transdisciplinary interventions.[9] Simultaneously, they gain a deeper understanding of how disciplinary-specific knowledge and skills can be used to facilitate participation-related goals in intervention. For example, a caregiver of a child with cerebral palsy reported that participation in mealtimes with her child is extremely difficult. A team assessment revealed that poor seating and positioning, difficulties in chewing as a result of increased oral tone, and difficulties in bringing food to the mouth with unadapted eating utensils, are some of the main reasons for difficulties within this routine. With the shared goal of improving participation in the mealtime routine, the PT may provide recommendations regarding adaptive seating for better postural control, which can also improve oral-motor movements by decreasing oral tone. In addition, the SLT may recommend experimenting with various food textures that may make it easier for the child to chew. The OT could also make recommendations for adapting eating utensils by providing advice and training on enlarging the handles of spoons or using a universal cuff to allow the child to manipulate a spoon better. The team then works together with the designated primary service provider or caseworker, by releasing the discipline-specific skills and strategies required to improve the shared goal of participation in the mealtime routine. The primary service provider, rather than all individual members of the team, works together with the caregiver and coaches them on the various discipline-specific strategies recommended by the team.

The environment for moving towards a transdisciplinary model of service delivery has been created due to the growing awareness internationally, as well as in South Africa, that the unidisciplinary model of training health professionals on its own is not fit for purpose in delivering an effective healthcare service.[16] After an initial workshop hosted in 2012 by the Academy of Science of South Africa (ASSAf), a subsequent multi-professional consensus study highlighted the need for interprofessional education and collaborative practice (IPECP) in the training of healthcare professionals.[16] IPECP occurs when two or more professions learn about, from, and with each other to enable effective collaboration and improve outcomes for individuals and the families they serve.[17]

IPECP and the transdisciplinary approach have the same underlying philosophy, namely shared learning across disciplines as well as collaborative teamwork. IPECP, however, does not stipulate the form that collaborative teamwork should take, although it is implied that it should be more collaborative than the siloed, multidisciplinary approach. The transdisciplinary approach takes the collaborative part of IPECP one step further by adding the components of integrated assessment and treatment goals that can be implemented by any of the team members (including community-based rehabilitation workers) acting as a primary service provider.

The ASSAf recommendation for IPECP is now slowly starting to be incorporated into the undergraduate training of rehabilitation professionals to train future members of interprofessional teams in South Africa who can more effectively address workforce constraints.[17,18] However, this approach to training may not be as pervasive across all institutions in the country that train healthcare professionals. Furthermore, the more integrated transdisciplinary philosophy of teamwork is not reflected in current undergraduate IPECP curricula in South Africa[17,18], although there is a focus on training on the International Classification of Disability, Functioning and Health in IPECP[14].

Despite these encouraging IPECP developments, my experience, as a university representative on the Gauteng Department of Health's ECI Forum[19] conducting workshops with community service rehabilitation professionals on collaborative ECI for children with disabilities, has shown that they still struggle to work collaboratively. The multidisciplinary approach to service delivery therefore still appears to be highly prevalent in the way that new graduates are trained.

I, therefore, propose recommendations for IPECP training in South Africa that may start to incorporate role release processes of the transdisciplinary approach[9] into undergraduate, postgraduate and continuing education programmes, especially as a significant percentage of the rehabilitation workforce may have not received any form of IPECP.

Role release or role transfer is where rehabilitation professionals, together with the family, share their basic expertise and allow the primary service provider to carry out the intervention plan with the child and family.[9] This concept is not unheard of in the traditional healthcare setting. During the COVID-19 pandemic, for example, many healthcare professionals in South Africa and elsewhere were trained to take on roles that were not part of their traditional scope of practice.[20] King and colleagues[9] outline key features of the role release process which build on each other and may be seen as the sum of several separate but related processes which should occur before a primary service provider can implement carryover of discipline-specific strategies with the child and family (Table 1). These processes, in hierarchical order, are role extension, role enrichment, role expansion, role exchange, role release and role support.[9] Table 1 outlines the key team member activities of each of these processes. It is suggested that for new trainees at an undergraduate level, role extension, role enrichment and role expansion training may be possible. Elements of role enrichment and role expansion already appear to be featured in some undergraduate IPE curricula[17] of rehabilitation professionals in South Africa[18]. Further role expansion at this level could already start to be incorporated into IPE practicums where students can be taught how to conduct an arena assessment.[9] The arena assessment is a defining feature of the transdisciplinary approach, in which each disciplinary team member assesses the child simultaneously, using both standardised measures and informal methods.[9] From a practical point of view, getting students from the rehabilitation disciplines together in the same space is quite feasible, especially as they tend to share some of the same practicum locations for their clinical blocks. During an arena assessment, one person from the rehabilitation team which consists of multiple disciplines, engages the child in a play activity. During the play routine, the other team members observe and collect information about different developmental areas, as well as the interrelatedness of these areas, together with the child's caregivers.

Table 1: The activities involved in the role release process at various levels of training

Type of process	Team member activities	Education level
Role extension	Team members are learning and expanding their discipline-specific clinical and practical skills in order to support children and families.	Undergraduate training / continuing education
Role enrichment	Team members start to learn and have a broader awareness about other disciplines of the rehabilitation team. Team members share information and best practices of their discipline as it relates to children.	
Role expansion	Team members share ideas and information to broaden their knowledge in order to make observations, judgements, and recommendations from multiple rehabilitation disciplinary perspectives, e.g. within an arena assessment and subsequent team meeting.	
Role exchange	Team members learn theories, methods, and procedures of other rehabilitation disciplines and begin to implement a few techniques from these disciplines with guidance.	Postgraduate training / continuing education
Role release	Individual rehabilitation team members are competent to use strategies and techniques from other disciplines without direct hands-on guidance.	
Role support	Team members continually advise, support and encourage each another, and in particular the primary service provider, especially when chosen interventions are more complex than most rehabilitation practitioners' training would have covered.	

The processes of role exchange, role release and role support are best trained at a postgraduate level or through continued education, as rehabilitation professionals need to have in-depth knowledge, experience and competence in their own disciplines before they can release discipline-specific skills over to a primary service provider. For a significant number of professionals who have received no IPE training, they will need to be trained in all processes and activities of role release through continuing education programmes. While the Gauteng Department of Health's ECI Forum[19] has started some of this training in ECI workshops, it may not always be targeting those who are most in need of it, for example, practitioners who were trained in multidisciplinary ways of working.

It should also be noted that not all graduate rehabilitation professionals may undertake postgraduate studies as they may not meet the eligibility criteria of universities. It is therefore an imperative to have more continued education programmes to train and support ECI rehabilitation teams in South Africa on the journey to becoming transdisciplinary, especially in rural settings.[3]

This is especially important as the current cohort of those trained at a postgraduate level on the transdisciplinary approach in South Africa anecdotally report that there are many systemic challenges in implementing transdisciplinary service delivery in their work contexts, despite their willingness to do so. These challenges include, but are not limited to, difficulty with managers who are unwilling to provide time for team training and coordination, as well as hospital systems that do not allow assessment and intervention to take place within the natural environments of the child, such as the home. These challenges necessitate not only more extensive research to be conducted with those trained in the transdisciplinary approach but also, for more evidence-based studies of transdisciplinary service delivery in South Africa, which still are lacking.[3]

Conclusion

The initial training time that teams need to become transdisciplinary should not be underestimated. However, the long-term benefits in terms of efficiency and effectiveness for already scarce professional rehabilitation resources[1] cannot be underestimated. The impending implementation of the NHI, which aims to address more equitable health services for all South Africans[6] even in the context of a scarce rehabilitation workforce, presents an important moment to pause and strategise on a sustainable way of delivering more efficient and effective services for children with disabilities, in line with a transdisciplinary approach.

Declarations

I have no competing interests to declare. AI was not used in the writing or editing of this manuscript.

References

1. Conradie T, Berner K, Louw Q. Describing the rehabilitation workforce capacity in the public sector of three rural provinces in South Africa. A cross-sectional study. Int J Environ Res Public Health. 2022;19(19), Art. #12176. https://doi.org/10.4102/sajp.v72i1.298
2. Ahmat A, Okoroafor SC, Kazanga I, Asamani JA, Millogo JJ, Illou MM, et al. The health workforce status in the WHO African Region: Findings of a cross-sectional study. BMJ Glob Health. 2022;7(Suppl 1), e008317; Reference 3. https://doi.org/10.1136/bmjgh-2021-008317
3. Visagie S, Swartz L. Rural South Africans' rehabilitation experiences: Case studies from the Northern Cape Province. S Afr J Physiother. 2016;72(1):1–8. https://doi.org/10.4102/sajp.v72i1.298
4. Health Professions Council of South Africa (HPCSA). Annual report 2022/2023. Pretoria: HPCSA; 2023. Available from: https://www.hpcsa.co.za/Content/upload/publications/annual_report/HPCSA_AR_2023.pdf
5. Padarath A, Moeti TL. South African health review 2022. Health systems recovery after COVID-19 [document on the Internet]. c2023 [cited 2024 May 24]. Available from: https://sahr.hst.org.za/article/87567-south-african-health-review-2022-health-systems-recovery-after-covid-19
6. Westwood A, Slemming W. Long-term health conditions in children: Towards comprehensive care [document on the Internet]. c2019 [cited 2024 May 24]. Available from: https://ci.uct.ac.za/sites/default/files/content_migration/health_uct_ac_za/533/files/CG2019%2520-%2520%25285%2529%2520Long%2520term%2520health%2520conditions%2520in%2520children.pdf
7. Samuels A, Slemming W, Balton S. Early childhood intervention in South Africa in relation to the developmental systems model. Infants Young Child. 2012;25(4):334–345. https://doi.org/10.1097/IYC.0b013e3182673e12
8. Castro-Kemp S, Samuels A. Working together: A review of cross-sector collaborative practices in provision for children with special educational needs and disabilities. Res Dev Disabil. 2022;120:Art. #104127. https://doi.org/10.1016/j.ridd.2021.104127
9. King G, Strachan D, Tucker M, Duwyn B, Desserud S, Shillington M. The application of a transdisciplinary model for early intervention services. Infants Young Child. 2009;22(3):211–223. https://doi.org/10.1097/IYC.0b013e3181abe1c3
10. Mahoney G, Wiggers B. The role of parents in early intervention. Implications for social work. Child Sch. 2007;29(1):7–15. https://doi.org/10.1093/cs/29.1.7
11. Ned L, Tiwari R, Hess-April L, Lorenzo T, Chikte U. A situational mapping overview of training programmes for community-based rehabilitation workers in southern Africa. Strategies for strengthening accessible rural rehabilitation practice. Front Public Health. 2020;8, Art. #569279. https://doi.org/10.3389/fpubh.2020.569279

12. South African Department of Health. Regulations defining the scope of the profession of speech-language therapy [document on the Internet]. c2017 [cited 2024 Jun 20]. Available from: https://www.hpcsa.co.za/Content/upload/professional_boards/slh/regulations/Regulations_defining_the_scope_of_the_profession_of_Speech_Language_Therapy.pdf

13. American Speech-Language-Hearing Association (ASHA). Scope of practice in speech-language pathology [document on the Internet]. c2016 [cited 2024 Jun 28]. Available from: https://www.asha.org/policy/sp2016-00343/

14. Filies GC, Müller J. The international classification of functioning, disability and health framework as a strategy to promote interprofessional collaboration during rural training in South Africa. In: Eretzen DV, Jacobs-Nzuzi Khuabi LAJ, Bardien F, editors. Transformation of learning and teaching in rehabilitation sciences: A case study from South Africa. Cape Town: AOSIS; 2022. p. 169–190. https://doi.org/10.4102/aosis.2022.BK357.08

15. Alant E, Mophosho M. Early childhood intervention. Web-based training for transformation. S Afr J High Educ. 2003;17(1):185–199. https://doi.org/10.4314/sajhe.v17i1.25208

16. Volmink J. Reconceptualising health professions education in South Africa. S Afr J Sci. 2018;114(7/8), Art. #a0281. https://doi.org/10.17159/sajs.2018/a0281

17. Pitout H, Barnard-Ashton P, Adams F, du Toit S. An African perspective on collaborative interprofessional curriculum evolvement. A qualitative reflection [preprint]. Research Square; 2022. https://doi.org/10.21203/rs.3.rs-2213687/v1

18. Maree C, Bresser P, Yazbek M, Engelbrecht L, Mostert K, Viviers C, et al. Designing interprofessional modules for undergraduate healthcare learners. Afr J Health Prof Educ. 2017;9(4):185–188. https://doi.org/10.7196/AJHPE.2017.v9i4.853

19. Balton S, Vallabhjee A, Burger E. Early childhood intervention. The Gauteng experience. S Afr Health Rev. 2020;2020(1):99–106.

20. Balton S, Pillay M, Armien R, Vallabhjee AL, Muller E, Heywood MJ, et al. Lived experiences of South African rehabilitation practitioners during coronavirus disease 2019. Afr J Disabil. 2024;13, Art. #1229. https://doi.org/10.4102/ajod.v13i0.1229

AUTHORS:
Robert Mash[1]
Jenny Nash[2]

AFFILIATIONS:
[1]Division of Family Medicine and Primary Care, Stellenbosch University, Cape Town, South Africa
[2]Eastern Cape Department of Health, Amathole District Clinical Specialist Team, East London, South Africa

CORRESPONDENCE TO:
Robert Mash

EMAIL:
rm@sun.ac.za

HOW TO CITE:
Mash R, Nash J. District health service delivery and the contribution of family physicians. S Afr J Sci. 2024;120(11/12), Art. #18712. https://doi.org/10.17159/sajs.2024/18712

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
primary health care, district hospitals, district health system, health reforms, family physicians

PUBLISHED:
30 August 2024

District health service delivery and the contribution of family physicians

Significance:

Most health care in South Africa takes place in the district health system. In the public sector, this includes primary health care and district hospitals. Although there have been improvements in health and health care, there are still inequities, many instances of poor quality, and weakness regarding community engagement and multisectoral action. Service delivery is currently challenged by budget cuts and loss of resources. Ongoing reforms are needed to improve performance and accommodate the introduction of national health insurance. The deployment of family physicians is an overlooked reform that can improve the model of care, quality, and resilience.

In this Commentary, we look at the current state of district health service delivery in South Africa and comment on the potential contribution of family physicians to improving delivery as part of future health reforms.

Understanding district health service delivery

Primary health care (PHC) is at the heart of district health services, with support from district hospitals. Governments around the world re-committed themselves in the 2018 Astana Declaration to strengthening PHC.[1] They stated that

> *strengthening PHC is the most inclusive, effective and efficient approach to enhance people's physical and mental health, as well as social well-being, and that PHC is a cornerstone of a sustainable health system for universal health coverage (UHC) and health-related Sustainable Development Goals.*

Following this statement, the World Health Organization (WHO) published an operational plan that defined the key components of PHC as integrated primary care services with essential public health functions, empowerment of people and communities, and multisectoral policy and action.[2] The operational plan was supported by a new conceptual and measurement framework for the health system through a PHC lens.[3] This framework defined the essential processes of service delivery as being the model of care, systems for improving the quality of care, and resilient health facilities and services. These processes determine the access, availability, and quality of service delivery. Ultimately, service delivery should enable universal health coverage (UHC), improved health status and health equity.

The term 'model of care' refers to the design, selection and planning of services.[3] What services should be offered in the community, in primary care facilities and at district hospitals? It also includes the organisation of services and how they are managed. In addition, it includes attention to community-based services, community linkages and engagement, and not just facility-based services.

Systems for improving quality of care should be organised in a systematic manner across all the services, measure the core functions of primary care, and monitor patient safety.[3] Sometimes this is referred to as clinical governance. Performance management systems should not just measure performance, but should also enable critical reflection, planning and implementation of interventions to improve quality and safety.[4] These processes should be cyclical and continuous to improve and maintain quality.

'Resilience of facilities and services' refers to their ability to prepare for, respond to and recover from shocks and challenges.[5] In recent times we have witnessed the need for resilience of district health services in the face of the COVID-19 pandemic[6], extreme climate events[7] and now cuts in the health budget[8].

What does district health service delivery look like?

In South Africa, at least 80% of the population is dependent on public sector services. In primary care the key service providers are nurses and nurse practitioners, sometimes supported by doctors. Services are provided by small clinics, often in rural areas, community day centres (open during office hours) and community health centres (open 24 hours). The community health centres often have a large multidisciplinary team that may include midwives, social workers, health promoters, dieticians, allied health professionals, dentists and pharmacists. Sometimes, community health centres have a midwife obstetric unit where people can deliver babies and an emergency centre open 24 hours a day. Often, additional services are provided as outreach on a periodic basis, for example, from psychologists, therapists or hospital-based specialists.

The public sector has also been employing ward-based PHC outreach teams that consist of community health workers and nurse coordinators.[9] These teams are allocated a delineated geographic area (often a municipal ward) and each community health worker has specific households for which they are responsible. Often, these teams extend the reach of the primary care facility into the community. They can follow up with patients or help link people back into care. Ideally, they should help the services shift the model of care from a focus on the patients in the facility to a focus on the health needs of the whole community served by that facility. This model of community-orientated primary care helps to provide more health promotion and to prevent or detect diseases earlier.[10] The approach can also identify and address the health needs of the community and some of the underlying social and environmental determinants. For example, the need for early childhood development centres or dealing with illegal dumping of hazardous waste. Community participation and stakeholder collaboration are essential ingredients.

At the same time as deploying community health workers, the public sector also created district clinical specialist teams to improve maternal and child health.[9] These district-level teams are meant to consist of a family physician, paediatrician, obstetrician, paediatric nurse, midwife, nurse practitioner and even an anaesthetist. They are responsible for improving the quality of maternal and child health services throughout the district. In many provinces there is a call for these district-level teams to have an expanded role to also engage with non-communicable diseases.

In the public sector there is a close relationship between the PHC platform and the next level of expertise at the district hospital. Small- and medium-sized district hospitals are found throughout the rural areas, and a few are also in our metropoles. Larger district hospitals in metropolitan areas are often organised along the lines of general hospital specialties – medicine, surgery, paediatrics, obstetrics and gynaecology. Most district hospitals, however, are generalist environments with wards for men, women, children, and maternity. There is usually an outpatient department and an emergency centre. Typically, the doctors in the district hospital provide outreach to the primary care platform. In several provinces, clinical associates also work in district hospitals. Sometimes there is a family physician at the district hospital who is a specialist in family medicine and provides support to the team.

The private sector looks after approximately 20% of the population and primary care is provided by general practitioners (GP). These GPs may work in solo or group practices and may work for themselves or for private health organisations. The private sector does not have the equivalent of district hospitals and tends to refer immediately to specialists and sub-specialists at the hospital. The GP practices do not usually have multidisciplinary teams, although other health professionals such as psychologists or physiotherapists may have their own practices in the vicinity.

Current state of district health service delivery in South Africa

The underlying social and environmental determinants of health have been worsening.[11] There is a desire to radically re-design the system with the introduction of national health insurance.[12] Contracting units for primary health care must still be clearly defined, but might centre around sub-districts and district hospitals. There is a consensus that health system re-design should enable a focus on population health and well-being, health promotion and disease prevention.[11]

There are many positive indicators in the South African health sector. The maternal mortality ratio (MMR) has decreased since 2021/2022 to 101/100 000 live births in 2022/2023.[13] Only the Western Cape (with an MMR of 62/100 000 live births) reached the Sustainable Development Goal (SDG) target of MMR <70/100 000 live births. The early neonatal death rate, which provides an indication of the quality of antenatal and intrapartum care, has continued to decrease over the past decade and is now 9.6/1000 live births.[13] There are currently 5 million people living with HIV receiving antiretroviral therapy in primary care, which makes up 34.9% of primary care spending.[13]

The Ideal Clinic and Ideal Hospital programmes are a precursor for the NHI. The programme commenced in 2015/2016, and, at that time, only 322 (9%) of the 3473 clinics in South Africa attained 'ideal clinic' status. This figure improved to 2706 (78%) clinics in 2023/2024, with 97% of clinics in KwaZulu-Natal, Mpumalanga and North West Provinces attaining ideal clinic status.[14] The Ideal Hospital programme commenced in 2018/2019, and of the 391 hospitals at the conception of the programme, only 8 (2%) attained ideal hospital status. In 2023/2024, 62 (16%) hospitals were classified as ideal. However, there are large variations between provinces, with 68% of hospitals in Gauteng achieving ideal hospital status, while none in Mpumalanga and the Northern Cape did.[14]

In September 2020, *The Lancet Global Health* Commission on High Quality Health Systems reported on the quality of health in low- and middle-income countries.[15] It noted that approximately 9 million lives are lost globally due to lack of quality of care, and about 60% of these deaths were amongst people who had managed to access health care. South Africa was one of six countries which hosted national consultations on the quality of health systems.[15] One of the concerns with the Ideal Clinic and Ideal Hospital programmes is that an ideal building infrastructure which contains the required medications, protocols and guidelines does not guarantee quality of care. Healthcare workers are required to implement these guidelines to provide high-quality health care, and communities need to have access to these facilities. In terms of the core functions of primary care, patients rate accessibility as poor.[16] In many communities there is no access to primary care in evenings or weekends and working people are disadvantaged. Appointment systems in areas of high utilisation may only provide access after 4–8 weeks. Many patients with primary care problems attend emergency centres to access care. Continuity of care is also rated poorly and people rarely see the same primary care provider.[16] There is still no electronic medical record and informational continuity is a challenge. Comprehensiveness, coordination, and cultural competence have been rated as stronger aspects of the system.[16]

The WHO sees community participation and empowerment as a central component of PHC.[2] In South Africa, patients have rated community orientation as poor.[16] Formal statutory structures should be supplemented by more informal and inclusive community health forums that can include civil society organisations and community leaders. There is also a need to move beyond engagement as a way of informing communities to a more participatory approach. The community-orientated primary care approach expects communities to participate in prioritising their health needs and planning action.[10]

The third pillar of PHC is multisectoral policy and action.[2] A focus on health and well-being requires health professionals to work with other sectors such as social services, education and police. Our ability to collaborate across sectors to address health issues is limited. Health professionals may lack skills in building effective collaborations with people from other sectors of society or government, may lack motivation and confidence to reach out beyond their focus on service delivery, and may not receive support from their managers or organisational environment. At a local level, the community-orientated primary care approach expects primary care providers to engage with a broad variety of stakeholders who can influence health in the catchment area.[10] Key issues such as violence prevention and mental health need an all-of-government approach.[11]

How can family physicians contribute to district health service delivery?

In our view, family physicians are one of the most underutilised solutions to some of the problems facing district-level service delivery.[17] Family physicians are qualified doctors who have spent a further 4 years of training to become specialists in family medicine. Training programmes are available at all the medical schools in South Africa, but since the speciality was created in 2007 only about 200 doctors have graduated.[18] Training posts are quite limited and human resource policy documents have misunderstood the role of family physicians.[19] Previous policies have seen family physicians as a sub-speciality of internal medicine or as specialists who should work at tertiary hospitals. Family physicians are trained to work independently at district hospitals and within primary care teams. Currently, only a third of the graduates have been retained in the public sector as posts are very limited.[20] In addition, 10% have emigrated and 11% have stopped practising medicine. Most of the family physicians have been employed in the Western Cape where the health system has committed to have family physicians at district hospitals and primary care facilities.[20] The South African Academy of Family Physicians has published a position statement that recommends a mid-term goal of one family physician at every district hospital, community health centre or sub-district (if there is no community health centre).[17] To achieve this modest goal we need at least another 400 family physicians, but at current rates of training it could take more than 20 years.

The contribution of family physicians to district health service delivery has been conceptualised as threefold.[17] Firstly, they contribute as clinicians and consultants to their healthcare teams. Secondly, they contribute through the capacity building and clinical training of those same teams. Lastly, they contribute through leadership of clinical

governance activities to improve the quality of care and patient safety. This contribution can be within community-orientated primary care or the district hospital. We discuss their contribution according to the three service delivery processes outlined by the WHO.

Family physicians and the model of care

Family physicians bring a higher level of expertise closer to the community.[21] In primary care, they can manage patients with complex conditions and multimorbidity who might otherwise need to be referred to secondary or tertiary level hospitals. They can supervise and support nurses and junior doctors within PHC. This improves the availability of comprehensive primary care to the community. Improving the level of confidence and trust in primary care may also reduce the tendency for people to bypass primary care and present themselves at the hospital.

At the district hospital, family physicians close skills gaps and enable the full package of care to be safely offered.[21] For example, they can improve the management of women in labour and reduce the incidence of maternal or neonatal catastrophes. Much of the litigation in the Department of Health is due to adverse maternal or neonatal outcomes. Likewise, they bring surgical and anaesthetic skills to the hospital and can perform operations such as amputations or emergency laparotomies.[22] This can save lives, make services more available and reduce the need for referral. In rural areas, referral to specialist care at distant hospitals is often difficult and delayed.

Family physicians are not trained as managers, but their leadership often strengthens the organisation and management of services.[23] All training programmes include a focus on the 'I-we-it' leadership model. This model looks at personal leadership style, values and development (the 'I'), leading through teams and relationships (the 'we') and engaging health systems and services (the 'it'). Family physicians bring a level of systems thinking and broad understanding because of their postgraduate training and an agency that helps teams to solve problems and perform better.[21] As leaders of the clinical team they can also advocate effectively for improvements such as equipment or service re-design. Many family physicians are employed as clinical managers.

Family physicians are also trained in the principles of community-orientated primary care and can thus assist with implementation[24], for example, improving the integration of facility-based and community-based team members or engaging with local stakeholders from social services. They may enable direct clinical support of community health worker teams or coordination between levels of care or with stakeholders in the community.

Family physicians and systems for improving quality

Family physicians are trained to lead clinical governance activities.[25] Such activities can include the development and implementation of evidence-based guidelines or protocols as well as audit and feedback on the technical quality of care for specific diseases or conditions. They can lead risk reduction strategies by reviewing unexpected mortality or morbidity and investigating patient safety incidents. They may also assist clinical teams to reflect on routinely collected data. For example, data on inappropriate antibiotic prescribing to reduce the chance of developing resistance, or data on laboratory tests to use resources more efficiently.[21] They may also assist with making a community diagnosis and identifying local health needs on the basis of data collected by community health workers. Their postgraduate training also equips them to participate in applied research projects to address research questions that are important for service delivery.[26] Recent questions addressed by a family physician research network include coordination of care with hospital outpatients[27] and exploring factors that influence the retention of medical officers[28]. They can also assist teams to interpret research evidence and decide on whether to change clinical practice.

Another way in which family physicians can improve quality is through clinical teaching and training.[25] Having a family physician in the team can enhance healthcare workers' confidence and motivation as they know someone is there to support them should they need help.[22] Family physicians are trained in a set of educational skills to provide constructive feedback to other members of the team and facilitate learning.[29] They can help develop an organisational culture of learning[30] that cascades down so that everyone is helping others learn and develop.[28] Family physicians also enable more formal clinical training by taking responsibility for registrars, interns, and medical and clinical associate students. All interns must now spend 6 months in family medicine and primary care, and most medical schools are moving towards greater exposure of medical students to PHC.

Family physicians and resilience of facilities and services

Currently, the public health sector in South Africa is facing massive budget cuts.[8] This translates into fewer healthcare workers, closure of beds, reductions in access to elective surgery, loss of locum staff, and erosion of staff morale. Many doctors who recently completed community service are struggling to find employment in the public sector. These cuts are a challenge to the resilience of the district-level health services to continue to offer care.

The most cost-effective part of the health system is PHC. For example, investment in community health workers can both improve health status and save money.[31] Family physicians are a cost-effective intervention as they strengthen district-level service delivery including PHC, reduce litigation, and enable more efficient use of resources.[17] Although they require the creation of specialist-level posts, the return on investment should more than justify the commitment.

Conclusions

South Africa has made huge strides in improving many health indicators, particularly by providing antiretroviral therapy in PHC. However, the increasing pressures from non-communicable diseases, trauma-related conditions, maternal and child health challenges, and a constrained fiscus, necessitate a new approach. Further quality improvement of PHC services and district hospitals is needed, as well as community engagement and multisectoral action. Family physicians can make an important contribution to strengthening the model of care, systems for improved quality, and the resilience of district-level service delivery.

Declarations

R.M. is the immediate past president of the South African Academy of Family Physicians. J.N. is the current secretary of the South African Academy of Family Physicians. We have no competing interests to declare. AI was not used in the preparation of this article. Both authors read and approved the final manuscript.

References

1. World Health Organization (WHO). Astana Declaration on primary health care 2018 [webpage on the Internet]. c2018 [cited 2018 Dec 18]. Available from: https://www.who.int/primary-health/conference-phc/declaration
2. World Health Organization (WHO). Operational framework for primary health care: Transforming vision into action. Geneva: WHO; 2020.
3. World Health Organization (WHO). Primary health care measurement framework and indicators: Monitoring health systems through a primary health care lens [webpage on the Internet]. c2022 [cited 2022 Mar 13]. Available from: https://www.who.int/publications/i/item/9789240044210
4. Munar W, Snilstveit B, Aranda LE, Biswas N, Baffour T, Stevenson J. Evidence gap map of performance measurement and management in primary healthcare systems in low-income and middle-income countries. BMJ Glob Health. 2019; 4(suppl8), e001451. https://doi.org/10.1136/bmjgh-2019-001451
5. World Health Organization (WHO). Operational framework for building climate resilient health systems. Geneva: WHO; 2015.
6. Ray S, Mash R. Innovation in primary health care responses to COVID-19 in sub-Saharan Africa. Prim Health Care Res Dev. 2021;22, e44. https://doi.org/10.1017/S1463423621000451
7. Naidoo K, Manyangadze T, Lokotola CL. Primary care disaster management for extreme weather events, South Africa. Afr J Prim Health Care Fam Med. 2022;14(1), Art. #3778. https://doi.org/10.4102/phcfm.v14i1.3778
8. Mash B. World Health Day: SA's public health sector facing crisis amid budget cuts [webpage on the Internet]. c2024 [cited 2024 Apr 07]. Available from: https://mg.co.za/thought-leader/opinion/2024-04-07-world-health-day-sas-public-health-sector-facing-crisis-amid-budget-cuts

9. Matsoso M, Fryatt R, Andrews G. The South African health reforms, 2009-2014: Moving towards universal coverage. Pretoria: Juta; 2015.

10. Mash B, Ray S, Essuman A, Burgueño E. Community-orientated primary care: A scoping review of different models, and their effectiveness and feasibility in sub-Saharan Africa. BMJ Glob Health. 2019;4(suppl8), e001489. https://doi.org/10.1136/bmjgh-2019-001489

11. South African Department of Health. Draft DHS strategy (2024–2030) executive summary booklet: Powering the health system from below. Pretoria: National Department of Health; 2024.

12. Government of South Africa. Act No. 20 of 2023: National health insurance [document on the Internet]. c2024 [cited 2024 May 17]. Available from: https://www.parliament.gov.za/storage/app/media/Acts/2023/Act_20_of_2023_National_Health_Insurance.pdf

13. Ndlovu N, Padarath A, editors. District Health Barometer 2022/23. Cape Town: Health Systems Trust; 2024.

14. South African Department of Health. Ideal Clinic South Africa: Snapshot of progress made [webpage on the Internet]. c2024 [cited 2024 May 17]. Available from: http://www.idealhealthfacility.org.za

15. Kruk M, Pate M. *The Lancet Global Health Commission* on high quality health systems 1 year on: Progress on a global imperative. Lancet Glob Health. 2020;8(1):E30–E32. https://doi.org/10.1016/S2214-109X(19)30485-1

16. Bresick G, von Pressentin KB, Mash R. Evaluating the performance of South African primary care: A cross-sectional descriptive survey. S Afr Fam Pract. 2019;61(3):109–116. https://doi.org/10.1080/20786190.2019.1596666

17. South African Academy of Family Physicians. The contribution of family physicians to district health services in South Africa: A national position paper by the South African Academy of Family Physicians. S Afr Fam Pract. 2022;64(1), Art. #a5473. https://doi.org/10.4102/safp.v64i1.5473

18. Tiwari R, Mash R, Karangwa I, Chikte U. A human resources for health analysis of registered family medicine specialists in South Africa: 2002–19. Fam Pract. 2021;38(2):88–94. https://doi.org/10.1093/fampra/cmaa084

19. South African Department of Health. 2030 Human Resources for health strategy: Investing in the health workforce for universal health coverage. Pretoria: National Department of Health; 2020.

20. Jacobs G, Mash R. The career pathways of new family physicians in South Africa from 2008 to 2022. S Afr Fam Pract. 2024;66(1), Art. #a5904. https://doi.org/10.4102/safp.v66i1.5904

21. Mash R. The contribution of family physicians to African health systems. Afr J Prim Health Care Fam Med. 2022;14(1), Art. #3651. https://doi.org/10.4102/phcfm.v14i1.3651

22. Hendriks H, Adeniji A, Jenkins L, Mash RJ. The contribution of family physicians to surgical capacity at district hospitals in South Africa. Afr J Prim Health Care Fam Med. 2021;13(1), Art. #3193. https://doi.org/10.4102/phcfm.v13i1.3193

23. Mash R, Blitz J, Malan Z, von Pressentin K. Leadership and governance: Learning outcomes and competencies required of the family physician in the district health system. S Afr Fam Pract. 2016;58(6):232–235. https://doi.org/10.1080/20786190.2016.1148338

24. Mash R, Gaede B, Hugo JF. The contribution of family physicians and primary care doctors to community-orientated primary care. S Afr Fam Pract. 2021;63(1), Art. #5281. https://doi.org/10.4102/safp.v63i1.5281

25. von Pressentin K, Mash R. Strengthening the district health system through family physicians. S Afr Health Rev. 2018;(1):33–39.

26. Mash R. Establishing family physician research networks in South Africa. S Afr Fam Pract. 2020;62(1), Art. #5216. https://doi.org/10.4102/safp.v62i1.5216

27. Mash R, Steyn H, Bello M, von Pressentin K, Rossouw L, Hendricks G, et al. The quality of feedback from outpatient departments at referral hospitals to the primary care providers in the Western Cape: A descriptive survey. S Afr Fam Pract. 2019;61(6):252–259. https://doi.org/10.1080/20786190.2019.1676021

28. Mash R, Viljoen W, Swartz S, Abbas M, Wagner L, Steyn H, et al. Factors influencing retention of medical officers in the district health services of the Western Cape, South Africa: An exploratory descriptive qualitative study. S Afr Fam Pract. 2022;64(1), Art. #a5467. https://doi.org/10.4102/safp.v64i1.5467

29. Mash R, Blitz J, Edwards J, Mowle S. Training of workplace-based clinical trainers in family medicine, South Africa: Before-and-after evaluation. Afr J Prim Health Care Fam Med. 2018;10(1), Art. #1589. https://doi.org/10.4102/phcfm.v10i1.1589

30. Mash B, Edwards J. Creating a learning environment in your practice or facility. S Afr Fam Pract. 2020;62(1), Art. #5166. https://doi.org/10.4102/safp.v62i1.5166

31. Daviaud E, Besada D. Saving lives, saving costs: Investment case for community health workers in South Africa. Cape Town: South African Medical Research Council; 2017.

AUTHORS:
Jacqueline Hoare[1]
Heidi Matisonn[2]

AFFILIATIONS:
[1]Department of Psychiatry and Mental Health and Neuroscience Institute, University of Cape Town, Cape Town, South Africa
[2]The Ethics Lab, Department of Medicine and Neuroscience Institute, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Jacqueline Hoare

EMAIL:
Jackie.hoare@uct.ac.za

HOW TO CITE:
Hoare J, Matisonn H. From the burden of disease to the disease of burden. S Afr J Sci. 2024;120(11/12), Art. #18716. https://doi.org/10.17159/sajs.2024/18716

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
moral suffering, healthcare workers, burnout, healthcare crisis, public health system

PUBLISHED:
30 August 2024

From the burden of disease to the disease of burden

Significance:

The burden of care that healthcare workers in South Africa shoulder as a result of the failings in service delivery in the public healthcare sector is having devastating effects on the healthcare workers themselves. Moral distress is increasingly being recognised as the primary occupational hazard for clinicians working in South Africa. But recognising the problem is not enough; we need to talk openly about it. Moral suffering needs to be discussed in medical curricula, professional development environments, and in our hospitals and clinics. We need to change the culture of 'suffering in silence' to one of 'supporting in community'.

In 2009, Mayosi et al.[1] described South Africa as being "in the midst of a profound health transition that is characterised by a quadruple burden of communicable, non-communicable, perinatal and maternal, and injury-related disorders". The so-called 'colliding epidemics' of HIV and tuberculosis, chronic illness and mental health, injury and violence, and maternal, neonatal, and child mortality, have been recognised to have "had substantial effects on health and well-being"[2]. In this Commentary, we draw attention, not to the patients who are burdened with these diseases, but to the (public) healthcare workers who attend to them. We argue that the burden they shoulder as a result of the failings in service delivery in the healthcare sector is similarly a disease, one which is reaching epidemic proportions and having devastating effects on healthcare workers and on the healthcare system in which they work.

Background

South Africa has a "fragmented, two-tiered and inequitable health system"[3]. A decade ago, approximately 30% of the doctors in the country served around 84% of the population who were dependent on the public healthcare system.[3] More recent estimates suggest that "less than 20% of our medical workforce is employed to serve approximately 75–80% of our population"[4]. In 2022, South Africa's doctor to patient ratio was 0.31 per 1000 – a decrease from 0.79 per 1000 in 2019. There is also a critical lack of nurses in the country: over 5000 nursing posts remain unfilled[5]; with only 22 090 nurses, a ratio of 1 nurse for every 2300 people, or 1 per 224 patients, and a significant number set to retire in the next 10 years, it is estimated that there will be a shortage of between 131 000 and 166 000 nurses by 2030.[6]

The infrastructure is no better off: a decade ago, Mayosi and Benatar pointed to the "state of crisis" that many of the state hospitals were in, "with much of the public healthcare infrastructure run down and dysfunctional as a result of underfunding, mismanagement, and neglect"[3]. If there has been a change in the physical environment of the public healthcare system, it has been for the worse: "clinics and health facilities are overcrowded and characterised by massive infrastructure backlogs"[4]. The 2024/2025 budget year of the Health Facility Infrastructure Grant saw an 11.7% cut – totalling ZAR1.2-billion – a cut which followed the previous year's 5.2% cut as a "cost containment measure."[4]

The 2024/2025 budget for the public health wage bill came in at ZAR174.6 billion, which is a 2.5% increase year-on-year, but if we take into account inflation, this figure translates to a decrease of 2.5%.[7] The total budget for health care – ZAR271.9 billion – also saw a 2.9% real cut once adjusted for inflation. What these numbers mean in practice is that while "spending per healthcare user increased nominally from R5180 in 2023/24 to R5243 in 2024/25, once adjusted for inflation the real spend per health care user was cut by R173 for each of the 51.9 million public healthcare users."[7]

These cuts, due in large part to many years of corruption and financial mismanagement at the level of the national government, need to be understood in the context of an already struggling healthcare system.[8] In 2023, the system was described as "sick"[9]; and in 2024, as facing an "existential crisis."[4]

Moral suffering in healthcare workers

Against this backdrop, and in the context of a society characterised by poverty, deep inequality and violence, the environment in which healthcare work is performed is deeply distressing in itself. Moral distress is increasingly, albeit slowly, being recognised as the primary occupational health hazard for clinicians working in South Africa.[10]

The term 'moral distress' originated in the field of health care, specifically nursing. It was initially conceptualised and defined by American ethicists Jameton and Boss in 1984 as a psychological and emotional response experienced by healthcare professionals when they believe they know the morally right course of action but are unable to act accordingly due to various constraints such as institutional policies, hierarchical structures, legal and ethical dilemmas, or conflicting values within a healthcare setting. Jameton later refined the concept, explaining moral distress as (a) the psychological distress of (b) being in a situation in which one is constrained from acting (c) on what one knows to be right.[11] It is the experience of being seriously compromised as a moral agent in practising in accordance with accepted professional – and personal – values and standards. To put it simply, moral distress is knowing what the right thing to do is but being (or feeling) unable to do it due to circumstances beyond one's control.

A related concept to moral distress is moral injury, defined in 1995 by Shay as "being present when i) there has been a betrayal of what is morally right, ii) by someone who holds the legitimate authority, and iii) in a high stakes

situation"[12]. While moral injury was originally studied in the context of Vietnam soldiers' experiences of PTSD, it has more recently been applied to the healthcare setting, specifically in the context of the COVID-19 pandemic, with its significant impacts on healthcare systems and providers. As Cartolovni et al.[13] argue, moral injury is "unique to those who bear witness to intense human suffering."

Individuals experiencing either moral distress or moral injury are generally at a higher risk of developing chronic physical and emotional exhaustion, as well as a sense of reduced accomplishment and autonomy, typically resulting from prolonged exposure to high levels of stress and impossible workloads that make excessive demands on energy, strength, or resources in the workplace. Work-related psychological distress in health workers may be associated with burnout and long-term adverse effects such as decreased quality of patient care, conflict with colleagues, cognitive impairment, substance use, suicide, poorer physical health, and leaving the healthcare profession.[14]

A host of studies have shown the physical and mental effects of moral distress and injury. They do not only negatively affect the healthcare provider, but directly and indirectly affect the efficiency and quality of healthcare delivery. Studies have found a causal relationship between moral distress and job performance, burnout and workplace attrition and ultimately, and most concerningly, patient care. In societies characterised by deep inequality and violence, such as ours in South Africa, the environment in which clinicians work is distressing in itself. This compounds the experience of moral distress in healthcare workers as resource shortages in the system combined with patients' very difficult personal circumstances leads to frequent exposure to distressing ethical situations.[15]

According to the Medical Protection Society survey released in October 2023, more than a third (35%) of healthcare practitioners in South Africa say their mental well-being is worse now than it was during the COVID-19 pandemic, and 39% of those surveyed said that the impact of burnout and exhaustion on patient safety was impacting their mental well-being. A quarter of the participants considered leaving the medical profession or retiring early due to mental well-being concerns and 22% said they planned to or considered leaving South Africa for those same reasons.[16] Other studies have "indicated that over 70% of young doctors working in primary care have burnout [from the] cumulative effects of mental and emotional stress, high workload in substandard facilities, and job instability."[4]

Care for the carers – what needs to be done?

How do we heal the healthcare system and its people?

First, we – all of us – need to hold our leaders to account: in May 2023, President Cyril Ramaphosa addressed participants at the Second Presidential Health Summit and acknowledged the need for "Government [to] invest more in training programmes for healthcare professionals and increase staffing levels to meet the population's needs"[7]. Yet there are now fewer healthcare workers than there were, as posts have been frozen in an attempt to deal with the budget cuts. All of us need to advocate for a halt to and reversal of austerity measures – because all of us are affected by them. To end the current crisis caused by national government, we need a reinvestment in public health urgently.

Second, we need leaders to listen to those on the ground such as the more than 16 academic heads of departments and nearly 1000 senior clinicians, nursing leaders and healthcare workers in the Western Cape, who voiced their concerns in an open letter[17] addressed to (then) Minister of Finance Enoch Godongwana, Western Cape Premier Alan Winde and the province's MEC for Finance Mireille Wenger. Penned by Ntobeko Ntusi and Lydia Cairncross, the letter highlighted the "devastating" impact that the severe budget cuts are having on the province's healthcare system.

While systemic change takes time and money, we can and must, in the interim and as a matter of urgency, recognise that the problem of moral suffering exists, is real, and, most importantly, is not shameful. This last point is critical: what exacerbates the experience of moral distress or poor mental well-being in healthcare workers is that there is a tendency for them to feel isolated and lonely in their suffering.[10] One contributing factor may be that, in the medical profession, vulnerability and uncertainty are seen as a sign of weakness, and this perception is reinforced by a culture of silence which convinces clinicians that their colleagues are successfully managing these stresses. A high rate of loneliness in medical settings has been correlated with poor work organisation, less managerial support, worse atmosphere in the team, and more irresponsible attitudes of colleagues. Loneliness in clinicians can be used to predict occupational burnout which in turn is correlated with high rates of attrition.[10] The connection and sense of belonging within the clinical team can provide the strongest protection against despair and loneliness. Each clinician needs to take responsibility for a change in medical culture. We as clinicians must create time for attending weekly peer support sessions, where we can, as clinical teams, explore parts of ourselves together, think and connect together within our unit/division/department, in a way that encourages and maintains psychological safety, social connections and a sense of belonging.[10]

But recognising the problem is not enough: we need to talk openly about it. Moral suffering needs to be discussed in medical and allied healthcare curricula, in professional development environments, and in our hospitals and clinics. We need to change the culture of health care from one of 'suffering in silence' to one of 'supporting in community'. Fawcett and Mullan argue that "fostering a practice culture where ethical issues and situations that give rise to moral distress can be openly discussed is important for mental health and wellbeing."[18]

Dealing with the problem requires skill – a particular kind of skill. We need to provide training for healthcare workers to develop ethical reasoning and critical thinking skills which are needed to have the confidence to discuss and make complex ethical decisions. Peer argues that "The skill to ethically reason is as important as technical and research-based skills and transcends all aspects of a profession. Just as technical and clinical skills mature with practice, so do ethical skills."[19] Schafer and Vieira suggest that ethics education "can help to promote the workers' ethical competence. Ethically competent professionals have greater skill in coping with the ethical questions posed by practice and, as a result of this, are better able to deal with the moral distress and its consequences."[20]

At the same time, we need to create spaces for healthcare providers to work together to find solutions. Khaghanizadeh et al.[21] have reported positive effects on moral reasoning, moral distress and moral sensitivity in nurses through group discussions. Clinical Ethics Forums offer healthcare workers the space to practise their reasoning skills in non-emergency settings and provide a psychologically safe environment where healthcare providers, particularly those from minority or underrepresented groups, feel valued and included, and where the climate encourages recognition of and support for each other's distress in making morally injurious clinical decisions in resource constrained healthcare facilities. Staff should have protected time to attend such forums which are a resource for developing ethical resilience in themselves, and in their clinical teams. This has been done before – at the height of the COVID-19 pandemic.[10] It is imperative that the system allows – and promotes – time for this again.

Lewis[22] claims that the most important way to prevent moral distress is to design an environment aimed at encouraging moral decision-making and empowering decision-makers to act. We also need to focus on creating support structures for complex clinical ethics decisions to be made in real time. Clinical Ethics Committees should be constituted; for smaller settings, these can be inter-institutional. Such committees function to promote, uphold, and respect the dignity and rights of patients and healthcare workers and can provide support and advice for navigating complicated ethical dilemmas. Moreover, they provide opportunities to build and strengthen a culture of collaboration in which the emphasis is on 'doing right' rather than 'being right'.

In addition, we need to find funding for the establishment of clinical ethics consultation services and/or clinical ethics posts embedded within hospitals.

Conclusion

Our proposed solutions signal a fundamental shift away from the idea of individualised support where the pathology is placed on the healthcare providers themselves who need to seek out, find time, and often pay for interventions. Stigma, shame, and isolation are substantial barriers to the success of individual-focused well-being efforts. In addition, individual-level interventions can produce toxic side effects of causing feelings of stress by implying it is an individual problem and self-imposed.

The individualised support approach discounts the role that the healthcare environment plays in causing distress. Sharma and Cotton have suggested at least one of the challenges of dealing with moral distress for healthcare workers in low- and middle-income countries is the absence of support mechanisms.[23] In such instances, the healthcare worker becomes the "second victim."

We have proposed a multi-pronged approach to dealing with the moral suffering experienced by healthcare workers, that includes: training for healthcare workers to develop critical thinking and ethical reasoning skills, Clinical Ethics Forums for promoting collaboration in problem-solving, and Clinical Ethics Committees for sharing decision-making responsibilities. These are not offered as an alternative to addressing the systemic problem – corruption, inefficiency, and maladministration – but as complements. We reiterate our concern that urgent attention to all these measures is needed. As with any disease, the longer you wait to intervene, the more difficult it is to cure. Unless we provide real support for healthcare workers whose moral suffering is already reaching epidemic proportions, the disease of burden will be fatal.

Acknowledgements

We thank Prof. Jantina de Vries of the EthicsLab for her review of a draft of the article.

Declarations

We have no competing interests to declare. We have no AI or LLM use to declare. Both authors read and approved the final manuscript.

References

1. Mayosi BM, Flisher AJ, Lalloo UG, Sitas F, Tollman SM, Bradshaw D. The burden of non-communicable diseases in South Africa. Lancet. 2009;374(9693):934–947. https://doi.org/10.1016/S0140-6736(09)61087-4
2. Achoki T, Sartorius B, Watkins D, Glenn SD, Kengne AP, Oni T, et al. Health trends, inequalities and opportunities in South Africa's provinces, 1990–2019: Findings from the Global Burden of Disease 2019 study. J Epidemiol Community Health. 2022;76(5):471–481. https://doi.org/10.1136/jech-2021-217480
3. Mayosi BM, Benatar SR. Health and health care in South Africa – 20 years after Mandela. N Engl J Med. 2014;371:1344–1353. https://doi.org/10.1056/NEJMsr1405012
4. Nash J, Ross AJ, Naidoo M, Ras T, Brits H, Mathew S. Health budget cuts will be paid for by the most vulnerable. S Afr Fam Pract. 2024;66(1), Art. #5934. https://doi.org/10.4102/safp.v66i1.5934
5. Lencoasa M, Brown D. Opinion: More austerity is the wrong medicine for our public health crisis. Spotlight. 2023 October 27 [cited 2024 Jun 26]. Available from: https://www.spotlightnsp.co.za/2023/10/27/opinion-more-austerity-is-the-wrong-medicine-for-our-public-health-crisis/
6. Businesstech. South Africa desperately needs more nurses [webpage on the Internet]. c2023 [cited 2024 Jun 26]. Available from: https://businesstech.co.za/news/business-opinion/687459/south-africa-desperately-needs-more-nurses/
7. Aneesa. South Africa will spend nearly R320 less per learner and R200 less per health service user in the coming year. Section 27. 2024 February 22 [cited 2024 Jun 24]. Available from: https://section27.org.za/2024/02/south-africa-will-spend-up-to-r320-less-per-learner-and-r200-less-per-health-service-user-in-the-coming-year/
8. Maphumulo WT, Bhengu BR. Challenges of quality improvement in the healthcare of South Africa post-apartheid: A critical review. Curationis. 2019;42(1), Art. #a1901. https://doi.org/10.4102/curationis.v42i1.1901
9. Huisman B. Women in health: "No one taught me how to treat a sick system," says leading young Professor. Spotlight. 2023 August 11 [cited 2024 Jun 27]. Available from: https://www.spotlightnsp.co.za/2023/08/11/women-in-health-no-one-taught-me-how-to-treat-a-sick-system-says-leading-young-professor/
10. Hoare J. Provision of mental health care to healthcare workers during COVID-19: A call for the practice of vulnerability. S Afr J Sci. 2022;118(5/6), Art. #13904. https://doi.org/10.17159/sajs.2022/13904
11. Jameton A. What moral distress in nursing history could suggest about the future of health care. AMA J Ethics. 2017;19(6):617–628. https://doi.org/10.1001/journalofethics.2017.19.6.mhst1-1706
12. Shay J. The birth of tragedy – out of the needs of democracy. Didaskalia. 1995;2(2).
13. Cartolovini A, Stolt M, Scott PA, Suhonen R. Moral injury in healthcare professionals: A scoping review and discussion. Nurs Ethics. 2021;28(5):590–602. https://doi.org/10.1177/0969733020966776
14. Hoare J. The power of connected clinical teams: From loneliness to belonging. Philos Ethics Humanit Med. 2023;18(1):16. https://doi.org/10.1186/s13010-023-00143-7
15. Matisonn H, De Vries J, Hoare J. Capturing moral distress as a global phenomenon in healthcare. Am J Bioethics. 2023;23(4):82–84. https://doi.org/10.1080/15265161.2023.2186520
16. Medical Protection. Survey shines a light on mental wellbeing of doctors in South Africa [webpage on the Internet]. c2023 [cited 2024 Jun 26]. Available from: https://www.medicalprotection.org/southafrica/casebook-and-resources/medicolegal-articles-and-features/view/survey-shines-a-light-on-mental-wellbeing-of-doctors-in-south-africa
17. Metelerkamp T. Western Cape health professionals make united appeal for 'catastrophic budget cuts' to be halted. Daily Maverick. 2024 February 06 [cited 2024 Jun 27]. Available from: https://www.dailymaverick.co.za/article/2024-02-06-western-cape-health-professionals-make-united-appeal-for-catastrophic-budget-cuts-to-be-halted/
18. Fawcett A, Mullan S. Managing moral distress in practice. In Practice. 2018;40(1):34–36. https://doi.org/10.1136/inp.j5124
19. Peer K. Ethics education: Preventing moral distress and empathy decline in sports medicine practice. Int J Athl Ther Train. 2016;22(1):47–52. https://doi.org/10.1123/ijatt.2016-0024
20. Schafer R, Vieiera M. Ethical competence as a coping resource for moral distress in nursing text context nursing. Florianópolis. 2015;24(2):563–573. https://doi.org/10.1590/0104-07072015001032014
21. Khaghanizadeh M, Koohi A, Ebadi A, Azimin AE. The effect and comparison of training in ethical decision-making through lectures and group discussions on moral reasoning, moral distress and moral sensitivity in nurses: A clinical randomized controlled trial. BMC Med Ethics. 2023;24:58. https://doi.org/10.1186/s12910-023-00938-5
22. James Madison University. Meeting the challenge of moral distress [webpage on the Internet]. c2015 [cited 2024 Jun 26]. Available from: https://www.jmu.edu/news/2015/09/lewis-erica-discusses-moral-distress.shtml
23. Sharma D, Cotton M. We care for patients, but who cares for us? Trop Doct. 2021;51(2):141. https://doi.org/10.1177/00494755211007280