Advancements in environmental monitoring: Integrating digital tools for enhanced MiniSASS assessment and regional diatom index calculation

RCJ Koen

orcid.org/ 0000-0002-1583-2972

Thesis accepted in fulfilment of the requirements for the degree *Doctor of Philosophy in Science with Environmental Sciences* at the North-West University

Promoter: Prof JC Taylor

Graduation: May 2026

ACKNOWLEDGEMENTS

First and foremost, I give thanks to God, whose strength and guidance have sustained me throughout the completion of this study.

I extend my sincere gratitude to Prof. Jonathan Taylor for his exceptional mentorship, insightful advice, and consistent support. His guidance has been instrumental in shaping the direction and quality of this work.

My heartfelt thanks go to my parents, whose unwavering support and belief in my abilities made this opportunity possible.

I am deeply grateful to Mr. Rico Koen for his invaluable contributions in developing the digital diatom index tools and the miniSASS mobile application. His expertise and assistance have greatly enriched this study.

To Miss Danélle Steyn, thank you for your kind words, encouragement, and support throughout this journey.

I would also like to acknowledge the following institutions for their essential roles in supporting this work:

CGIAR, for providing the funding necessary to complete the miniSASS project.

GroundTruth, for their meaningful contributions to the development and implementation of the miniSASS project.

COALTECH, for funding the development of the AMD index. Project No. E2021-1.

Finally, the digital tools developed and presented in this study are available from the author upon request.

ABSTRACT

Freshwater in South Africa is largely supplied by rivers. However, the availability of water is limited by a semi-arid climate, high evaporation and low precipitation rates, and El Niño–Southern Oscillation (ENSO) events that contribute to droughts and floods. Pollution from industrial, mining, agricultural, and domestic activities, including acid mine drainage, nutrient runoff, and organic waste, damages ecosystems and further reduces freshwater availability. To manage freshwater sustainably, continuous monitoring of water resources and public awareness are essential.

A holistic approach to managing water resources can be achieved using the tools of Integrated Water Resource Management (IWRM). Through the National Water Act (NWA) and the National Water Resource Strategy (NWRS), IWRM is implemented by Catchment Management Areas (CMA) within Water management Areas (WMA). Monitoring rivers, wetlands and estuaries is essential to inform sustainable water use, pollution control and ecological health. This is done, in part, through biomonitoring using diatoms (phytobenthos), aquatic macroinvertebrates and fish.

Biomonitoring provides a holistic view of water quality changes over time using biological indicators. Unlike chemical analysis, biomonitoring integrates the responses of biota in the determination of water quality, ecosystem health and habitat degradation. The phytobenthos and aquatic macroinvertebrate assemblages together provide an integrated indication of ecosystem resilience. Macroinvertebrates are excellent indicators of ecosystem functioning while diatoms indicate more specific water quality changes. Diatom monitoring is implemented through indices such as the Indice de Polluosensibilité Spécifique (IPS), the Generic Diatom Index (GDI), Trophic Diatom Index (TDI) and the Biological Diatom Index (BDI), whilst aquatic macroinvertebrates are implemented using the South African Scoring System (SASS5) and the mini Stream Assessment Scoring System (miniSASS). These indices are used as proxies for water quality that inform on pollution and water quality changes to, in turn, inform management of aquatic resources. Development of regional diatoms indices, as well as simplifying the available tools can help support water management in South Africa by increasing broader public awareness and providing experts and non-experts with newly revised tools.

Implementing digital tools that integrate diatom and macroinvertebrate indices will improve data collection and reduce data loss when communicating with online data repositories and cloud-storage. Additionally, incorporating training tools and machine learning techniques can increase identification accuracy, and in turn increase the reliability of data generated by these tools. Digital tools can bridge knowledge gaps and engage citizen scientists and experts to provide better information on water quality. By combining these tools, South Africans can stand as advocates for sustainable water management and aquatic ecosystem health.

This study aimed to develop a miniSASS mobile application that fully incorporates and improves data capturing for miniSASS surveys and incorporating machine learning to automatically identify macroinvertebrates to improve the applicability of miniSASS as a citizen science tool. Furthermore, this study aimed to develop new regional riverine diatom indices by using Weighted Averaging (WA), Generalised Logit Regression (GLR) and inferred knowledge to calculate diatom optimal environmental conditions and tolerance ranges, under nutrient content, organic content and ionic load. Additionally, a diatom index to detect Acid Mine Drainage (AMD) in wetlands surrounding coalmines is also included. The novel diatom indices are further incorporated into digital platforms that provide easy-to-use software for diatom index calculation and data generation without the need for OMNIDIA.

The miniSASS mobile application provides a holistic and modern tool to simplify and increase the efficiency of river health assessment. The mobile application includes a landing page with relevant information on how to use the app, a map page where users can explore monitoring sites and a sites creation page which houses the full capability of a miniSASS survey whilst including a newly designed digital classification key to improve identification and classification accuracy. Furthermore, the site creation page also incorporates a machine learning model trained on 13 000 images of macroinvertebrates to help alleviate the bottleneck of manual verification of site scores on the miniSASS website. An additional about page is included that provides tutorials, support and extra resources to educated users.

Riverine diatom indices were developed by incorporating optima and tolerances calculated using WA, for electrical conductivity (EC), dissolved inorganic nitrogen (DIN) and orthophosphate (PO_4^{3-}). Species optima were also inferred from expert knowledge related to nutrient levels, organic load and ionic composition. Indices were accordingly calculated for each parameter and combined into a final multimetric average index using weights. The GLR approach was tested but ultimately rejected due to the nature of the dataset. The available datasets were sparse and zero-inflated, with an infrequent occurrence of species along environmental gradients, ultimately making WA are more feasible approach. The calculated optima and tolerances for many species aligned with those present in the IPS, confirming the use of WA for optima and tolerance calculation The calculated indices correlated strongly with water quality and accurately reflected the ecological condition of the rivers used to create the indices. The indices calculated for the individual parameters reflect specific water quality changes more accurately, where the combined multimetric indices and the IPS reflect overall water quality best.

A dedicated AMD index was developed by calculating species optima for sulphate, EC, pH, alkalinity and chloride, as these parameters characterize AMD disturbance in wetlands. These optima were integrated with AMD and osmotic tolerance values together with life-form categories (motile, attached and tube-forming) to calculate a final multimetric index score. The index successfully distinguished between AMD disturbed and non-disturbed sites with pH and sulphate emerging as the strongest environmental drivers. The correlation of the index scores with water quality was confirmed using bootstrapping with 1000 iterations of index scores. Four key taxa were identified as early indicators of AMD disturbance in wetlands. *Nitzschia capitellata* and *Frustulia crassinervia* indicate AMD impacted sites with low pH and high sulphate. *Amphora veneta* and *Craticula molestiformis* indicate AMD free sites, correlating with increased chloride and alkalinity.

The diatom indices developed in this study were fully integrated into digital tools that support efficient and accessible water-quality assessment of rivers and wetlands. A Diatom Indexer was created to house the riverine indices calculated using WA and knowledge inferred, as well as the widely used IPS index. The software generates an interpretable, illustrative graph that allows users to track changes in water quality across

sites and provides autecological information on the species included in the index calculation together with the scores for each index. A separate AMD Indexer was created to process a species matrix and produce a summary of AMD index scores. Both digital tools use a separate standard list as a reference for calculated optima and tolerances, and contain the AMD tolerances, osmotic tolerances and life-form scores for AMD index calculation. Together these tools enhance data generation, data throughput and knowledge dissemination, by providing free to use- user friendly software that streamlines the efficiently of diatom index calculation.

The miniSASS mobile application helps bridge the gap between public awareness of water quality and scientific knowledge. Coupled with the additional newly developed diatom index tools biomonitoring is made simpler by providing visual, accessible as easy-to-use tools to empower citizens and community members to engage and participate in river health assessments. Simultaneously, these digital tools provide a reliable and scientific method that can support decision makers and management practitioners. Therefore, by improving public awareness and bolstering water quality monitoring, these tools help South Africa move closer to achieving SDG6 - ensuring the availability and sustainable management of clean water and sanitation for all by 2030.

Keywords: Digital biomonitoring tools, Diatom indices, Acid Mine Drainage, MiniSASS, Machine Learning.

TABLE OF CONTENTS

List of Figures

List of Tables

optima and tolerances; ABS – absolute optima and tolerances; SP – sum product calculation; MIN/MAX – minimum/maximum transformation; LIN – Linear regression).

Table 10: Pearson correlations of selected indices (all correlations are significant with $p < 0.05$)

Table 11: Ecological classification of sites as determined by the IPS and other calculated indices in this study.

Abbreviations

AI	Artificial Inteligence
AMD	Acid Mine Drainage
ANN	Artificial Neural Network
ASPT	Average Score Per Taxon
ASTERICS	AQEM/STAR Ecological River Classification System
ASV	Abundance Sensitivity Value
AV	Abundance Value
AQEM	Assessment System for the Ecological Quality of Streams and Rivers Throughout Europe using Benthic Macroinvertebrates
BDI	Biological Diatom Index
BGCMA	Breede-Gouritz Catchment Management Area
BIODISCOVER	Biological specimens Described, Identified, Sorted, Counted and Observed using Vision-Enabled Robotics
BOD	Biological Oxygen Demand
CA	Correspondance Analysis
CCA	Cononical Correspondance Analysis
CEC	Commission of the European Communities
CEMAGREF	Centre National du Machinisme Agricole, du Génie Rural, des Eaux et des Forêts
CMA	Catchment Management Areas
CNN	Convolutional Neural Network
COD	Chemical Oxygen Demand

DIC	Differential Interference Contrast
DIN	Dissolved Inorganic Nitrogen
DO	Dissolved oxygen
DWA	Department of Water Affairs
DWAF	Department of Water Affairs and Forestry
DWS	Department of Water and Sanitation
EC	Electrical Conductivity
ENSO	El Niño–Southern Oscillation
EPS	Extracellular Polymeric Substances
FAQ	Frequently Asked Questions
GCIS	Government Communication and Information System
GDI	Generic Diatom Index
GLR	Generalised Logit Regression
GMSA	Global System for Mobile Communications Association
GSM	Gravel, Sand and Mud
GWP	Global Water Partnership
H_2O_2	Hydrogen Peroxide
HCl	Hydrochloric Acid
IHI	Index of Habitat Integrity
IPS	Indice de Polluosensibilité Spécifique
IUCMA	Inkomati-Usuthu Catchment Management Area
IWRW	Integrated Water Resource Management
KBI	Knowledge Based Index
KBI-A	Knowledge Based Index Average
KBI-I	Knowledge Based Index for Ionic load
KBI-N	Knowledge Based Index for Nutrients
KBI-O	Knowledge Based Index for Organic load

$KMnO_4$	Potassium Permanganate
KZN	Kwa-Zulu Natal
miniSASS	Mini Stream Assessment Scoring System
MIRAI	Macroinvertebrate Response Assessment Index
ML	Machine Learning
MMI	Multimetric Index
NAEHMP	National Aquatic Ecosystem Health and Monitoring Programme
NEsMP	National Estuaries Monitoring Programme
NMMP	National Microbial Monitoring Programme
NTMP	National Toxicity Monitoring Programme
NWA	National Water Act
NWMP	National Wetland Monitoring Programme
NWRS	National Water Resource Strategy
ORASECOM	Orange-Senqu River Commission study
PO_4^{3-}	Orthophosphate
PTV	Pollution Tolerant Valves
RDA	Redundancy Analysis
REMP	River Eco-status Monitoring Programme
RHP	River Health Programme
RQIS	Resource Quality Information Services
RQO	Resource Quality Objective
SADI	South African Diatom Index
SANBI	South African National Biodiversity Institute
SASS	South African Scoring System
SDG	Standard Development Goal
SIC	Stones In Current
SOC	Stones Out of Current

STAR	Standardisation of River Classifications
TDI	Trophic Diatom Index
VEG	Vegetation
VEGRAI	Vegetation Response Assessment Index
WA	Weighted Averaging
WAI	Weighted Average Index
WAI-AVG	Weighted Average Index Average
WAI-DIN	Weighted Average Index for Dissolved Inorganic Nitrogen
WAI-EC	Weighted Average Index for Electrical Conductivity
WAI-OP	Weighted Average Index for Orthophosphate
WA-PLS	Weighted Averaging using Partial Least Squares
WESSA	Wildlife and Environment Society of South Africa
WQ	Weighted Quantiles
WWTF	Waste Water Treatment Facility

CHAPTER 1: INTRODUCTION

1.1) Overview of freshwater supply and -pollution in South Africa.

Freshwater is one of the most important resources needed to sustain life. In South Africa, rivers are the main source of freshwater supply - there are virtually no natural lakes to exploit as a source of freshwater (Dallas & Day, 2004). Consequently, rivers are dammed to increase water supply for industrial, mining, agricultural, and domestic use. The climate of South Africa is semi-arid, and the annual precipitation rate is exceeded by the annual evaporation rate (DWAF, 1986). Therefore, the freshwater supply in South Africa is under additional strain due to climatic factors that reduce its availability. Additionally, the El Niño Southern Occislation (ENSO) phases also greatly influence the weather and rainfall across South Africa (Nhesvure, 2020), during El Niño phases, rainfall over most of the country is reduced and the risk of drought increases (Dieppois *et al*., 2015; Hoell *et al*., 2016). During La Niña, higher than normal rainfall occurs over the interior of the country leading to increased risk of flooding (Dieppois *et al*., 2015; Nicholson & Selato, 2000).

Flooding and drought, as hydrological extremes, have devastating effects on water availability and water quality (Death *et al*., 2015; Owolabi & Belle, 2023; Qiu *et al*., 2023). During drought, solutes in water become more concentrated, water temperatures rise and toxic substances like heavy metals become more readily available. Nutrient concentrations also increase leading to eutrophication (Van Vliet & Zwolsman, 2008). Although droughts may offer a minor advantage by containing contaminants locally and preventing their spread to surrounding waters, their overall negative impacts far outweigh any potential benefits (Qiu *et al*., 2023). Flooding can introduce harmful toxins and organic matter as well as nutrients into water bodies thereby altering the water chemistry and physical structure of rivers (Ching *et al*., 2015), however, flooding also removes sediment from rivers which often contain harmful substances, therefore increasing the water quality (Talbot *et al*., 2018). The impact of flooding and drought on water availability and quality is a dynamic and complex area of research (Fasihi *et al*., 2021). Flooding and drought also have devastating socio-economic impacts such as decreased food production during

drought and physical damage to property and alteration of geomorphology during flooding (FAO, 2018; Balgah *et al*., 2023).

ENSO phases are becoming increasingly difficult to predict and the duration and intensity of El Niño and La Niña phases can vary, creating increased uncertainty of the impacts of future La Niña and El Niño events, and planning for the effects elicited by such events becomes increasingly difficult (Zheng *et al*., 2022; Alizadeh, 2022).

Furthermore, greater pressure is placed on freshwater supply due to a rapidly growing human population that increases the scale of industrial, agricultural, and domestic practices that use freshwater to sustain such a growing population (Kock, 2017; Mnisi, 2020). South Africa is also an underdeveloped country and many of its citizens have limited or no access to potable clean water which necessitates the need for improved water supply networks (Edokpayi *et al*., 2018).

The supply of freshwater is further decreased by pollution invertedly created by industrial, mining, agricultural and domestic practices as well as land-use changes and afforestation (Dallas & Day, 2004; Musingafi, 2014). Iron and steel production produce hazardous chemical waste that makes its way into surrounding water bodies by either being directly discharged as effluent, or indirectly through seepage from slag heaps and landfills. The effluent discharge and seepage from steel and iron production contains high concentrations of suspended solids, metals, cyanide and oils (Hu *et al*., 2020; Garg Singh, 2022; Choudhury *et al*., 2023).

Petrochemical manufacturing that produces fuels by converting coals and gas use harmful chemicals such as benzene, xylene, toluene in production. Effluent from petrochemical manufacturing discharge contains toxic phenols, hydrocarbons and sulphates and if not correctly disposed of flows into surrounding wetlands and rivers causing detrimental effects to aquatic life (Stefanakis, 2021; Mohammadi *et al*., 2020; Varjani *et al*., 2020) Additionally, harmful air emissions of sulphur dioxide and nitrogen oxides contribute to air pollution which propagates through water cycles and cause acid rain, which acidified soil, weakens crop resilstance and interefers with soil microbial communities (Li *et al*., 2020; Chen *et al*., 2024). Automotive and cement manufacturing

also contribute to water pollution with effluent high in heavy metals, detergents and solvents (Rodrigues & Joekes, 2011; Ansari *et al*., 2013).

Mining, also considered an industrial activity, either being chemical or geological, creates a phenomenon known as acid mine drainage (AMD) (Musingafi, 2014). AMD is characterized by having a high concentration of dissolved metals, dissolved salts and high concentrations of sulphate and a low pH (Akcil & Koldas, 2006; McCarthy, 2011; Musingafi, 2014). This phenomenon occurs when old and abandoned gold/platinum and coal mines fill with rainwater and groundwater which eventually seeps into the environment, or when active mines use surrounding water bodies like wetlands, small lakes and rivers to discharge effluent rich in dissolved heavy metals (Akcil & Koldas, 2006; McCarthy, 2011). Some of the effects of AMD on human and aquatic life are the contamination of drinking water and the disruption of natural aquatic ecosystems by reducing the reproduction of aquatic macrophytes, fish and microbial communties (Reddick *et al*., 2018; Munyai *et al*., 2021; Windisch *et al*., 2025). Additionally, the mobilisation of heavy metals is increased because of the low pH associated with AMD, and so the dispersion of metals into surafce and groundwater in increased (Humphries *et al*., 2017) Wetlands are especially susceptible to the effects of AMD since many wetlands have been destroyed for mining and power generation, and many wetlands still surround mines. Wetlands serve important ecological functions such as containment of high nutrients loads and organic matter, purification of freshwater and barriers for flood protection (Humphries *et al*., 2017). Freshwater contaminated by AMD and used for agricultural practice has adverse effects by decreasing the fertility of soil that reduces crop yield, additionally the low acidity and presence of heavy metals can contaminate crops directly and cause mutations in the growth and reproductive cycle of such crops (Musingafi, 2014).

Agricultural practices focus on the production of food needed to sustain a population. Crop yield is increased by regular irrigation and fertiliser application (Sosibo *et al*., 2017; Lam *et al*., 2023). The semi-arid climate of South Africa already decreases the freshwater supply for irrigation and places additional stress on rivers to supply water for irrigation. Fertilisers used for agriculture contain high concentrations of nitrogen and phosphorous compounds which stimulate the growth of terrestrial and aquatic plants and algae (Mudaly

& van der Laan, 2020; Lukhele & Msagati, 2024). Runoff from agriculture enters nearby water sources as many farms are often situated near rivers for constant freshwater supply. Effluent containing high concentrations of nitrogen and phosphorous stimulates the growth of nuisance algae and can often lead to toxic algal blooms and eutrophication (Mudaly & van der Laan, 2020; Jwaideh *et al*., 2022; Lukhele & Msagati, 2024). Eutrophication directly influences freshwater systems causing an imbalance in food web interactions between aquatic organisms and by decreasing the aesthetic value of water through alteration in the smell, taste and the visual presentation of water (Watson, 2004; Oberholster & Ashton, 2008). Cyanobacterial blooms can lead to the death of aquatic invertebrates and fish due to a low dissolved oxygen (DO) concentration in water and an increase in the chemical and biological oxygen demand (COD and BOD respectively), in addition to the toxic neurotoxic and hepatotoxic compounds released (Codd *et al*., 1997; Chatterjee & More, 2023). Agriculture also increases the ionic load in rivers due to salinization through water abstraction and land use practices (Williams, 1999; Palmer & Coleman, 2004; Thorslund *et al*., 2021).

Furthermore, organic pollution decreases the availability and use of freshwater. Much of South Africa's population lives in rural and informal settlements that have little or no sewerage systems to remove waste produced by these communities (Armitage *et* al., 2009; Amoah *et al*., 2020). This sewerage enters nearby aquatic systems and contains high concentrations of organic matter and human waste. Organic pollution of freshwater systems also increases the BOD and COD within aquatic systems by releasing long carbon chain compounds that are difficult to break down, these chains require much oxygen to break down and therefore the BOD and COD is further increased (Dallas & Day, 2004).

The waste products and effluents released by the abovementioned practices reduce the freshwater supply of South African rivers. This places immense pressure on WWTFs to clean wastewater and make it fit for domestic, industrial use and agricultural use. Unfortunately, many of the WWTFs in South Africa fail to meet the required limits for constituents in effluents released (Herbig, 2019; Graham *et al*., 2025). Nearly half of WWTFs in South Africa fail to treat wastewater properly and many municipalities are in a critical state. The failure of WWTF's to treat wastewater properly stems from

mismanagement of plants and the unfortunate presence of corruption and nepotism (Herbig, 2019). Additionally, there is little enforcement of legislation and no criminal sanctions in place to ensure the proper functioning of WWTFs.

The Department of Water and Sanitation (DWS), previously called the Department of Water Affairs and Forestry (DWAF) defines water quality in its guidelines by its chemical, physical, and biological characteristics that determine its fitness for a variety of uses and for the protection of the health and integrity of aquatic ecosystems (DWAF, 1996a; DWAF, 1996b). This statement together with the abovementioned information shows the importance of continued monitoring of the water quality of freshwater aquatic ecosystems. Additionally, the dissemination of this information to communities and public of South Africa is important to both inform and engage the communities to become aware of such issues and to become advocates for the correct management of freshwater resources in South Africa.

1.2) Integrated Water Resource Management (IWRM) in South Africa

1.2.1) Background of IWRM in south Africa as implemented through National Resource Management Strategies (NWRS)

IWRM is a global ideology on the management of water resources. It is a holistic approach that aims at integrating the management of the water systems within the overarching political and socioeconomic framework (Claassen, 2013). It deals with different world views and as such navigates different societal and institutional beliefs, which can make its integration complex. IWRM is defined as a process that maximizes social and economic welfare through coordinated management and development of water, land and other related resources in such a manner that does not compromise the sustainability of vital ecosystems (GWP, 2000). In essence, IWRM aims at integrating the quality and quantity of all water resources within river basins to different usage sectors within the broader framework of national development planning (Claassen, 2013).

In 1998, South Africa promulgated the National Water Act, which is regarded by many countries as one of the most progressive and comprehensive water management legislations in the world. However, the implementation of the NWA in south Africa has been weak (Schreiner, 2013). The NWA is the primary legal framework for water

management in the country and mandates the control, protection, use, development and conservation of water resources (NWA, 1998). The National Water Resource Strategy (NWRS) is the official implementation of IWRM under the NWA (DWS, 2023). The NWRS is the backbone for IWRM in South Africa and includes ecological monitoring, planning, licencing and conservation at catchment level through catchment management agencies (CMAs). CMAs manage water resources within one or multiple Water Management Areas (WMAs), they are responsible for integrating IWRM at catchment level (DWS, 2023). Originally, nineteen CMAs were envisioned in the first NWRS, however, this was reduced to nine established CMAs in the second NWRS with the purpose to improve efficiency and address capacity constraints (DWA, 2013). As of 2020 only two of the proposed nine CMAs have been created; the Breede-Gouritz and Inkomati-Usuthu CMAs. The latest draft of the third NWRS released in 2023 further reduces the number of proposed CMAs to six and shifts the focus to national coordination and regional decentralisation, given the capacity exists (DWS, 2023). The proposed NWRS3 aims at more realistic planning strategies and reduced fragmentation.

1.2.2) Water Management Areas (WMA) in South Africa.

The implementation of IWRM as part of the NWRS, is done in WMAs. This section gives an overview of the WMAs in South Africa (Figure 1) and the industrial, agricultural and domestic practices within each area.

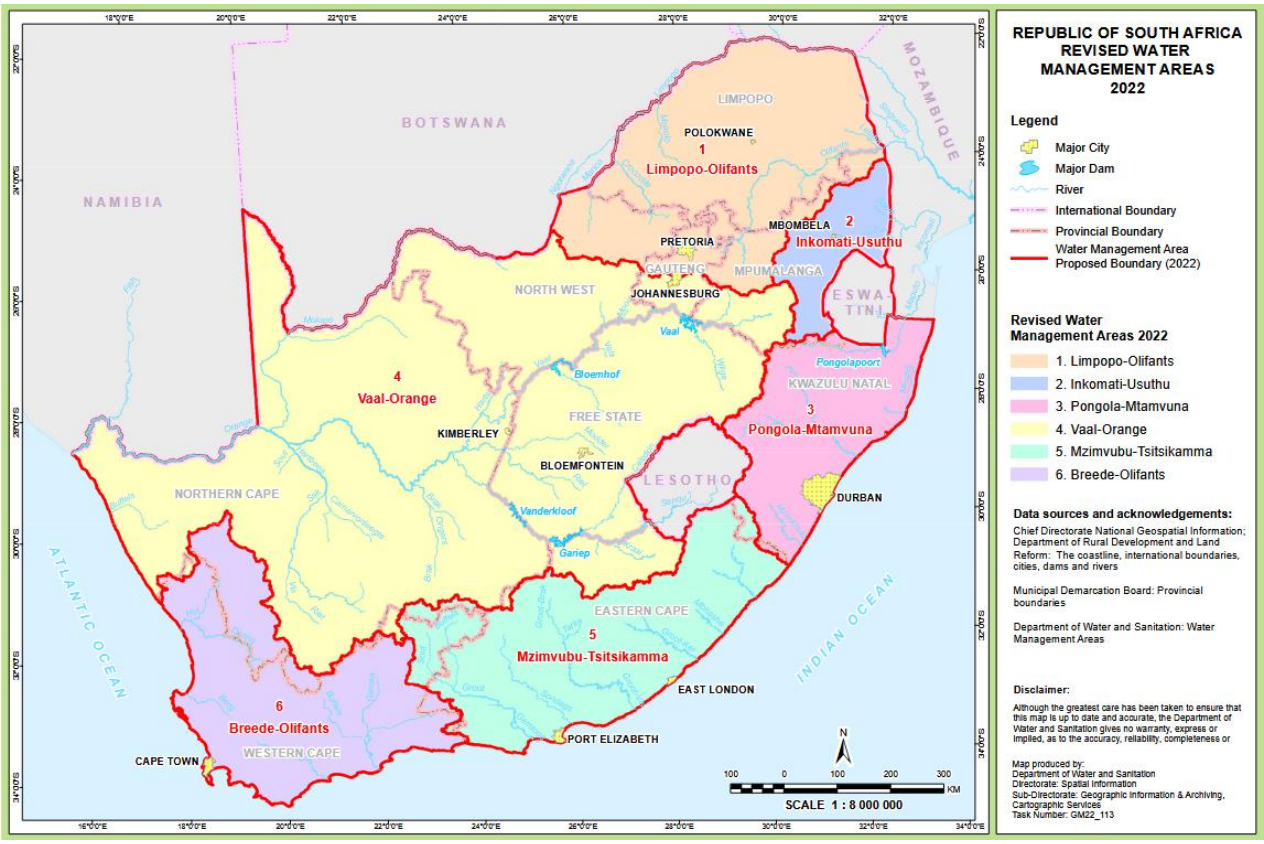


Figure 1: Map of the current WMAs in South Africa as of 2022 (DWS, 2023).

1.2.2.1) Limpopo-Olifants WMA (WMA01)

The Limpopo-Olifants WMA spans the northern part South Africa and falls within the larger Limpopo basin. The Limpopo basin covers an area of approximately 413 000 km^2 and spans four countries – South Africa, Botswana, Zimbabwe and Mozambique (Gründemann *et al*., 2018). Nearly half of the Limpopo basin (45 %) is situated in South Africa with the main tributaries of the Limpopo River in this region being the Crocodile, Marico, Matlabas, Mokolo, Lephala, Mogalakwena, Sand, Nzhelele, Luvuvhu, Olifants, Letaba, Steelpoort and Shingwedzi Rivers (Sitoe & Qwist-Hoffman, 2013). Together these rivers form the Limpopo-Olifants WMA which supplies water to most of Gauteng, Limpopo and parts of Mpumalanga. Three river catchments are in the Limpopo-Olifants

WMA namely the Crocodile-Marico catchment, Limpopo catchment and the Olifants catchment.

The Crocodile and Marico rivers together form the Crocodile (west)-Marico catchment. The Crocodile and Marico rivers supply Gauteng and the northern part of the North West province with water to support industrial, agricultural and domestic practices in these areas (DWS, 2016a). A growing population and increased urban development have increased the pressure on these rivers as the surrounding industrial and agricultural practices increase in size and intensity to sustain such a growing population (DWS, 2023). Domestic water use also increases and consequently the volume of sewage discharge also increases. With increased discharge from industry, agriculture and domestic activities the water quality of the Crocodile and Marico rivers has declined in the past decades (du Preez *et al*., 2018). Increased effluent requires more frequent water treatment and consequently increases the cost of maintaining water quality for domestic use and ecological reserve (Mitchell *et al*., 2014). Examples of ecological disturbance are the Roodekoppjes and Hartbeespoort dams. The nutrient loads in these dams are extremely high and causes ecological degradation and increases the cost of water purification for domestic use (Carroll & Curtis, 2021; Mnyango *et al*., 2025). Nuisance plants and toxic algal blooms are frequent in these dams, particularly in the Hartbeespoort dam.

In the Limpopo catchment, the main rivers are the Matlabas, Mokolo, Lephala, Mogalakwena, Sand and Nzhelele rivers. These rivers supply water to most of the Limpopo province and are dammed along the river reaches. These dams serve multiple purposes that include providing water for irrigation, mining, livestock and domestic practices (Lombaard *et al*., 2015). The main water use of these rivers is irrigation of croplands and livestock practices which accounts for approximately 75% of the water use, coal and platinum mining accounts for approximately 16% while domestic water use account for the remaining 9% (Lombaard *et al*., 2015). Domestic water use mainly depends on groundwater supply, which accounts for 40% of the volume within the Limpopo catchment, approximately 70% of domestic and small industrial water supply relies on groundwater discharge.

The Olifants Catchment catchment includes the Olifants River and its main tributaries. In the upper catchment the Wilge and Elands rivers form tributaries of the Oifants River, in the middle catchment the Steelpoort and Blyde rivers form tributaries and in the lower catchment the Ga-Selati, Letaba and Klasserie rivers form tributaries (DWS, 2022a). The Olifants River originates in Mpumalanga and flows north-east through Limpopo and eventually into Mozambique where it joins the larger Limpopo River.

The main land-use practices on the Olifants river include mining, power generation, domestic use and irrigated crop farming. The upper catchment supports coal mining, the middle catchment supports platinum mining, and the lower catchments supports copper and phosphate mining. Agriculture and irrigation account for 57% of the water use in the catchment, power generation accounts for 19% and domestic, industrial and mining operations account for an additional 19% (Kloos & Tsegai, 2009). The water quality of the Olifants river catchment is impacted by agriculture but more severely by the effects of AMD originating from coal, platinum, copper and phosphate mining (DWA, 2011; Dabrowski *et al*., 2015).

1.2.2.2) Inkomati-Usuthu WMA (WMA02)

The Inkomati-Usuthu WMA falls within the Inkomati River basin. This basin spans three countries (South Africa, Eswatini and Mozambique) and covers an area of approximately 36 500 km^2 within South Africa (IUCMA, 2023). The major rivers within this WMA are the Crocodile River (east), the Komati River, the Sabie River, the Sand River, which flows into the Sabie river. The Lomati River which flows into the Komati river is also included in this WMA. The main land-use practices in the Inkomati-Usuthu WMA are agriculture and forestry as well as industry and power generation. This WMA also supports domestic use and inter-basin transfers from South Africa to Eswatini (IUCMA, 2024).

The Inkomati-Usuthu CMA (IUCMA) oversees water management in the Inkomati and Usuthu river basins. The purpose of the IUCMA is to monitor river health through water quality monitoring, managing pollution and regulating water use (IUCMA, 2024). The CMA also oversees water allocation reform and engages with stakeholders to correctly manage the water use within these basins.

Agriculture and forestry accounts for 52% of the water use. Industry (paper mills) and mining account for 1% of the water use, domestic activities account for 5%, and 6% is accounted for by Cross-border transfers. Additionally, 23% of the water in the area is allocated as an ecological reserve (IUCMA, 2020).

1.2.2.3) Pongola-Mtamvuna WMA (WMA03)

The main rivers within this WMA are the Pongola and Mtamvuna rivers. The Pongola River, being the largest in this area, originates in northern Kwa-Zulu Natal and flows south-east into Mozambique. The Pongolapoort dam is one of the impoundments on this river and is one of the largest dams in South Africa with a volume of 2,445 million m^3 (Jaganyi *et al*., 2009; Moodley *et al*., 2021). The Mtamvuna River forms the border between Kwa-Zulu Natal and the eastern cape and flows into the Maputo River in Mozambique and forms part of the Maputo River basin. The irrigation sector in the Pongola catchment is the largest water use secotr, requiring approximately 213 million m^3/annum. This irrigation is mostly for commercial forestry, sugarcane production as well as fruit and vegetable cultivation (DWAF, 2004). The Pongolapoort dam supports local tourism and the Pongola floodplain, downstream of the Pongolapoort dam, support the local economy (Jaganyi *et al*., 2009).

1.2.2.4) Vaal-Orange WMA (WMA04)

The Orange River basin is the largest in Southern Africa and spans approximately 896,368 km^2. The basin spans four countries (Namibia, Botswana, Lesotho and South Africa), with most of the basin being in South Africa (64.2 %) (de Loë, 2009).

The Orange River originates in Lesotho, where it is known as the Senqu river. The Orange River flows Southwest and drains in the Atlantic Ocean and forms the border between South Africa and Namibia. The main tributaries of the Orange River are the Vaal, Fish, Caledon, Molopo and Nossob rivers (DWS, 2014). The Vaal River is one of the most the important river in South Africa as it supports much of the industrial and agricultural practices that sustain the economy, and contributes 36% of the total runoff of the Orange River (DWS, 2014). The Vaal River basin spans approximately 240 128 km^2 and flows through five provinces (Mpumalanga, Gauteng, Free state, North West and Norther-Cape) (Hendriks & Rossouw, 2009; DWS, 2022b). The Vaal River supports the Vaalharts irrigation scheme, which is one of the largest single irrigations schemes in the World,

irrigating 39 820 ha (DWS, 2022c). The lower Vaal supports 80 000 ha of land irrigation along the Vaal, Rietspruit, Harts and Douglas rivers (Barnard *et al*., 2012). Additionally, the Vaal River supplies water to the greater part of Johannesburg metro and most of the Free state (DWS, 2014). The Vaal River is also one of the most polluted rivers in South Africa due to effluent originating from domestic, agricultural and industrial activities (DWS, 2016b). Coal and gold mining contribute to pollution in the Vaal River through AMD and heavy metal contamination whilst domestic and agricultural practices release raw sewage, organic material and fertiliser into the river (McCarthy, 2011; DWS, 2016b).

1.2.2.5) Mzimvubu-Tsitsikamma WMA (WMA05)

This WMA lies between within the eastern cape, between Kwa-Zulu Natal and the Western Cape. The main rivers in this area are the Mzimvubu, Mthatha, Buffalo, Keiskamma, Sundays, Great Fish, Gamtoos and Swartkops rivers (Mulangaphuma & Jovanovic, 2025). The main land-use practices in this WMA are agriculture and forestry, domestic water supply as well as industry. Agriculture includes mostly irrigation for crops and comprises approximately 60% of the water use within this WMA (Mulangaphuma & Jovanovic, 2025). Forestry is prominent in this area and contributes to water pollution and alters ecological streamflow balances. Domestic water supply accounts for 17% of the water in this WMA and contributes greatly to water pollution since many of the WWTF are dysfunctional. The main impacts on the rivers in this area are nutrient pollution from surrounding agricultural practices and decreased streamflow due to exotic tree plantations (Mulangaphuma & Jovanovic, 2025).

1.2.2.6) Breede-Olifants WMA (WMA06)

This WMA lies mostly within the Western Cape and some parts within the Northern and Eastern Cape. Major rivers systems in this area are the Breede and the Gouritz River systems, the Gouritz River system include the Groot River, Gamka River and the Olifants River. The Breede-Gouritz CMA (BGCMA) oversees this WMA (BGCMA, 2023). They are responsible for coordinating river health monitoring and pollution control, to implement water reform and monitors illegal water abstraction aiming at water conservation, climate adaptation strategies and drought response (BGCMA, 2023). Most of the water in this area is allocated to agriculture, farming for fruits, vegetables and trees, and is approximately 95% of the water use in this area. The remaining water supply is allocated

to domestic use and industry. Industry includes tourism and aquaculture as well as agricultural processing and small-scale mining (BGCMA, 2023).

1.2.3) Resource Quality Information Services (RQIS)

In addition to the Catchment Management Agencies (CMAs) and Water Management Areas (WMAs), Resource Quality Information Services (RQIS) functions as the technical and scientific branch of the Department of Water and Sanitation (DWS). It provides essential data and assessment tools for national water quality monitoring and plays a key role in achieving the ecological monitoring goals of the National Water Resource Strategy (NWRS) (DWAF, 2004).

RQIS provides data to the DWS through the National Aquatic Ecosystem Health and Monitoring Programme (NAEHMP). Components of the NAEHMP includes the River Eco-status Monitoring Programme (REMP), previously known as the River Health Programme (RHP), The National Estuaries Monitoring Programme (NEsMP), the National Wetland Monitoring Programme (NWMP), the National Toxicity Monitoring Programme (NTMP) and the National Microbial Monitoring Programme (NMMP) (DWAF, 2004; DWS, 2023). Together these programmes broaden the water monitoring scope to include, rivers, estuaries, wetlands, toxicants and microbial communities within aquatic ecosystems to provide a more complete holistic overview of the state of aquatic ecosystems. This aims at improving data consistency and improved constitutional coordination and decision-making across multiple water resource programmes (DWAF, 2004; DWAF, 2007; Mangadze *et al*., 2019).

The NAEHMP employs different monitoring techniques to determine the ecological state of aquatic ecosystem, either using biological indicator groups or habitat assesment, or a combination of techniques. Diatom assesmblages, aquatic macroinvertebrate communities and fish assemblages form an integral part of the REMP, NEsMP and the NWMP (Wilkinson *et al*., 2016). Diatoms are used in the Indice de Polluosensibilité Spécifique (IPS), the Trophic Diatom Index (TDI), the Biological Diatom Index (BDI) and the Generic Diatom Index (GDI) (Coste in CEMAGREF, 1982; Kelly & Whitton, 1995; Coste & Ayphassorho, 1991) . Aquatic macroinverterbates are used in the South African Scoring System version 5 (SASS5), the Macroinvertebrate Response Asssessment Index

(MIRAI) and the mini Stream Assessment Scoring System (miniSASS) (Dickens & Graham, 2002; Graham *et al*., 2004; Thirion, 2007). Fish are used in the Fish Response Assessment Index (FRAI) (Kleynhans, 2007). These groups are used in different capcities in the REMP, NEsMP and the NWMP, and together they provide important insight into the overall ecosystem health of an aquatic system as well as any specific pollutants and toxicants within the system. Furthermore, habitat inegrity is determined through the Vegetation Response Assessment Index (VEGRAI) and the Index of Habitat Integrity (IHI) (Kleynhans *et al*., 2007; Kleynhans *et al*., 2008). These techniques are used together to provide a more holistic view of aquatic ecosystem health and function and have been implemented more widely within the context of water monitoring in South Africa. Diatom, aquatic macroinvertebrate and fish indices provide important information on the ecological condition of rivers, wetlands, and estuaries (Taylor *et al*., 2005; Harrison & Whitfield, 2006; Ollis *et al*., 2006, Malan *et al*., 2015, Adams *et al*., 2020). This information can be communicated to the Department of Water and Sanitation (DWS) and Catchment Management Agencies (CMAs) to support water resource management decisions under the National Water Act (NWA) and the National Water Resource Strategy (NWRS). By monitoring water quality and ecosystem health, these indices align with the achievement of Sustainable Development Goal (SDG) 6, particularly SDG 6.3 (improving water quality) and SDG 6.6 (protecting and restoring water-related ecosystems) (DWS, 2020).

1.3) Biological monitoring in South Africa

1.3.1) Implementation of diatoms and aquatic macroinvertebrates in the monitoring framework

Determining the integrated water quality of aquatic ecosystems is largely done by means of biological monitroing or biomonitoring. Biomonitoring uses biological indicator organisms that are constantly present within aquatic ecosystems and allows for an integrated view of water quality in contrast to chemical analysis which only provides a snap-shot image of current conditions (DWAF, 1996; Wepener, 2008).

In South Africa, diatoms and aquatic macroinvertebrates are incorporated as important biomonitoring tools within the River EcoStatus Monitoring Programme (REMP). This framework is structured around biological, habitat and chemical assessments used to

determine ecological classification of rivers and set Resource Quality Objectives (RQOs) (Feio *et al*., 2021). Aquatic macroinvertebrates are a crucial part of the REMP and are used as bioindicators in the South African Scoring System version 5 (SASS5), the mini Stream Assessment Scoring System (miniSASS) and the Macroinvertebrate Response Assessment Index (MIRAI). SASS5 is the primary biomonitoring tool applied in the REMP and the most widely used biological assessment (DWAF, 2008; Feio *et al*., 2021). MiniSASS serves primarily as a citizen science tool for early pollution detection and community engagement but has limitations as a biomonitoring tool in terms of precision, accuracy, and reliability (Graham and Taylor, 2018). The MIRAI assess habitat requirements for macroinvertebrates and are designed to assess river health through physical properties and habitats (Thirion, 2008; DWS, 2016c).

Diatoms are not as widely used as aquatic macroinvertebrates due to a lack of capacity and baseline information (Dalu & Froneman, 2016). Diatoms are therefore not seen as replacements for aquatic macroinvertebrates but rather as a supplementary tool, however, diatoms have seen a growing use in the monitoring framework as bioindicators (Taylor *et al*., 2004; de la Rey, 2008; Dalu & Froneman, 2016; Mangadze *et al*., 2018). The Indice de Polluosensibilité Spécifique (IPS), the Trophic Diatom Index (TDI), the Biological Diatom Index (BDI) and the Generic Diatom Index (GDI) serve are the main indices that incorporate diatoms as biological indicators and requires specialized computer software (OMNIDIA) to calculate, however, the indices do provide more specific information on environmental impacts than other rapid methods and can be used to identify specific sources of pollution (Leicointe *et al*., 1993).

1.3.2) Diatoms and aquatic macroinvertebrates as indicator groups

Biological indicators are those organisms that are considered integral in food web relations and connections and adhere to specific criteria that make them useful (Gerhardt, 2002). In South Africa the most common indicator organisms are diatoms, aquatic macroinvertebrates and fish, only diatoms and aquatic macroinvertebrates are considered in the present study. Each of these groups have advantages and disadvantages which necessitate the combined use of these organisms/groups in biomonitoring to obtain the most accurate determination of water quality.

Diatoms are unicellular algae that are ubiquitous in all aquatic ecosystems and some terrestrial ecosystems. Diatoms occurring in terrestrial ecosystems are seldom used for biomonitoring, however, recent studies have emphasised the importance of soil diatoms in ecolgoical assessments to imorve soil management, environmental monitoring and pollution assessment strategies (Yogeshwaran *et al*., 2025). Aquatic diatoms specifically epilithic diatoms are reliable indicators for specific water quality conditions and are excellent at indicating organic pollution, trophic condition and ionic loads in aquatic ecosystems (Taylor *et al*., 2007a; Dalu & Froneman, 2016). Furthermore, environmental conditions largely determine the diatom taxa regardless of geographic location, the structure of the diatom community is determined by other factors such as niche availability, physical structure and community succession in response to stressors (Dalu *et al*., 2014; Dalu & Froneman, 2016; Dalu *et al*., 2022). Diatoms can be regionally and globally comparable, which is a distinct advantage over other bioindicator organisms, although some debate exists on the validity of this statement due to gaps present in taxonomic identification. Diatoms have short generation times which allows for water quality monitoring over extended periods of time. Diatoms indicate short term changes (3-4 weeks) in water quality because they respond quickly, reliably and predictively to a wide range of pollutants (Bate *et al*., 2002). Additionally, diatoms are primary producers in aquatic systems and represent a large portion of the edible algae that higher trophic levels consume for food and energy (Round *et al*., 1990). Therefore, they form an integral part of food web function within aquatic systems. Diatoms possess a silica impregnated cell wall called a frustule which does not decompose and may accumulate in the sediment where they are preserved (rivers, lake and wetland), and as such can be used to infer past environmental conditions (Battarbee *et al*., 2001; Smol, 2008).

The disadvantages of using diatoms in biomonitoring is the relatively complex preparation and analysis of samples and the microscopy techniques necessary to identify and count diatom cells (Round *et al*., 1990; Taylor *et al*., 2007a). Sampling diatoms is an easy task, however, before counting and identifying diatom taxa, samples are digested in acid to remove all organic material to clean diatom frustules (Taylor *et al*., 2007b). This is crucial for species identification since the patterns formed in diatom frustules are species specific and are needed for identification below genus level. Microscopy techniques (light

microscopy and electron microscopy) can be a tedious task to undertake as it is difficult to identify diatoms to species level without appropriate guides or literature (Round *et al*., 1990; Taylor *et al*, 2007c; Smol & Stoermer, 2010). Therefore, using diatoms as biomonitoring tools requires laboratory analysis and advanced microscopy which gives delayed results, however, their advantages make such results highly valuable.

Aquatic macroinvertebrates are found in most aquatic ecosystems and respond to the physical and chemical characteristics of aquatic systems. They represent many taxonomic groups including insects (diptera, coleoptera, trichoptera, ephemeroptera, odonata, megaloptera and plecoptera), bivalves, gastropods, annelids and crustaceans. Consequently, aquatic macroinvertebrates represent many connections within aquatic food webs and are representative of overall ecosystems functioning in contrast to diatoms which indicate only specific environmental changes (Johnson *et al*., 1993; Dickens & Graham, 2002; Dallas & Day, 2004). The use of aquatic macroinvertebrates in biomonitoring has certain advantages and disadvantages. Sampling for aquatic macroinvertebrates is relatively easy. A net is used to catch species within three biotypes (aquatic vegetation, stones in and out of current, and gravel sand and mud) after which they are identified and counted. Unlike counting and identifying diatoms, macroinvertebrate groups/families are counted and identified *in situ* because they are visible to the naked eye (Dickens & Graham, 2002). Additionally, the distinct morphological diversity between groups makes identifying aquatic macroinvertebrate families relatively easy for an experienced technician with the aid of a guide. Aquatic macroinvertebrates serve as an important food source for fish and other aquatic vertebrates and therefore also form an integral part of aquatic food webs as do the diatoms (Johnson *et al*., 1993; Dallas & Day, 2004). Aquatic macroinvertebrates are more mobile than diatoms but less mobile than fish and therefore also represent current environmental conditions where they are found.

Aquatic macroinvertebrates have seasonal occurrence within aquatic systems and their diversity within aquatic systems is dependant on season. Autumn commonly yields a greater diversity of aquatic macroinvertebrates than spring and summer and consequently the method using aquatic macroinvertebrates will be influenced by seasonal dependence and scores will differ due to the difference in diversity of organisms counted

and sampled (Dickens & Graham, 2002). In addition to seasonal dependence, the presence of aquatic macroinvertebrates is controlled by favourable and unfavourable physical conditions. Biotope availability will determine the presence of aquatic macroinvertebrate groups, systems with higher diversity of biotopes will yield a greater score than systems with low biotope diversity, although this does indicate lower ecosystem resisiliance, it does not necessarily indicate on poor water quality (Dallas, 2007). The presence of floods and droughts can influence the diversity of aquatic macroinvertebrates as well as specific biotopes being either retained or removed from the system (Lake, 2000). When analysing and interpreting results, these conditions must be considered.

Using diatoms and macroinvertebrates in tandem is advantageous since the advantages and drawbacks of each group is balanced by the other. The collective use of diatoms and aquatic macroinvertebrates for biomonitoring of aquatic systems yield results that explain the specific environmental conditions (pollution types, trophic state and ionic loads) as well as the overall ecosystem functioning and resilience of an aquatic system to environmental change, therefore, giving a more complete overall picture of water quality within aquatic ecosystems.

1.3.3) Indices incorporating diatoms and aquatic macroinvertebrates

To use diatoms and aquatic macroinvertebrates to infer water quality within an aquatic system, requires the use of an index. An index, in the context of biomonitoring, is a mathematical formula generally incorporating sensitivity and tolerance values of taxa and weighing those values against the abundance of taxa found in a sample (Smol & Stoermer, 2010). Within a sample with a certain level water quality degradation, taxa with their optimum close to that level will be highest in abundance and will contribute more to the index value than those that are rare (Hawkins *et al*., 2000). The calculation of the index incorporates the sensitivity and tolerance values of all taxa counted, weighted against their abundance, to produce a single value that reflects the water quality of a given site within an aquatic system. The sensitivity value or optimal environmental condition reflects a taxon’s ability to compete well against other taxa and become high in abundance within the system under a specific concentration of an environmental variable

(Taylor *et al*., 2007a). Tolerance values are used in diatom indices and reflect their ability to tolerate changes under the environmental variable, meaning as the values for the variable increase or decrease, the taxon can withstand such changes and remain present in the system (Johnson *et al*., 1993). The sensitivity and tolerance values are inferred from environmental variables including but not limited to trophic level, ionic load and organic pollution that are important factors influencing water quality degradation (Hawkins *et al*., 2000). For example, the sensitivity and tolerance values for taxa to ionic load (high or low) will reflect the water quality as influenced by the ionic load.

Many indices for diatoms and aquatic macroinvertebrates have been developed, but only certain indices have value within South Africa for the purpose of water quality monitoring as well as for use in citizen science. The origin and history of such indices will be explained further in the following chapters. The most important diatom indices used in South Africa are the Indice de Polluosensibilité Spécifique (IPS), the Generic Diatom Index (GDI) and the Trophic diatom index (TDI). The IPS index incorporates over 13 000 taxa identified to species/variety and forma level with sensitivity and tolerance values ascribed to those taxa. The IPS reflects water quality as influenced by trophic state, ionic load and organic pollution and serves as an important contribution to water quality monitoring in South Africa since most water quality degradation has origin in one or a combination of these factors (Coste in CEMAGREF, 1982; Griffin *et al*., 2014). The IPS has been adapted for use in South Africa by including endemic species and forming the South African Diatom Index (SADI) (Taylor *et al*., 2007a). The GDI originally included 174 taxa at genus level and is less specific than the IPS; however, the difference in taxonomic resolution does not offer a significant statistical difference in results (Coste & Ayphassorho, 1991; Szczepocka *et al*., 2014). This is noteworthy considering that not all species within a genus share similar sensitivity and tolerances to environmental variables, making species-specific identification important for accuracy. The GDI is derived from the IPS and therefore also reflects water quality changes as influenced by trophic state, ionic load and organic pollutants. The TDI is not a general water quality index but does incorporate a % pollution tolerant valves (PTV) value that represents the number of taxa tolerant to organic pollution and should be used as an auxiliary tool for decision making for water quality management, particularly wastewater treatment works effluent (Kelly &

Whitton, 1995). The diatom indices mentioned above were not developed in South Africa, however, the application of these indices was tested by Taylor (2004) and Taylor et al. (2007a) and found to be successful for indicating water quality degradation in South Africa and have been tested and employed henceforth and up to the present.

Concerning aquatic macroinvertebrates, two indices stand out as important biomonitoring tools in South Africa namely the South African Scoring System (SASS) and miniSASS (mini–Stream Assessment Scoring System). In contrast to diatom indices these were developed in South Africa. SASS was originally developed by Chutter (1994) and has since been upgraded to the 5th version of SASS (SASS5) by Dickens and Graham (2002) and has been the biotic index of choice within the country for monitoring water quality. SASS5 incorporates 91 families of aquatic macroinvertebrates that represent larger taxonomic groups (Dickens & Graham, 2002). Each family is ascribed a sensitivity or optimum environmental condition value that is weighed by the abundance of that family sampled within three distinct biotopes. The score for each family is summed to produce a SASS score after which a simple average calculation is performed to produce an ASPT value (Average Sensitivity Per Taxon) by dividing the SASS score by the number of taxa present. Complimentary to SASS, miniSASS and was developed by Graham *et al*. (2004) as a citizen science tool for early pollution detection. miniSASS incorporates larger taxonomic groups compared to SASS5. Because of the lower taxonomic resolution used in miniSASS, the technique is less sensitive to fine-scale changes in the community composition in comparison to SASS5, furthermore the accuracy of miniSASS depends largely on correct identification and correct sampling technique. However, miniSASS is advantageous in that the public and communities of South Africa can use the index with relative ease and contribute to mapping the water quality state of aquatic systems in South Africa (Graham *et al.,* 2004; Taylor *et al*., 2022).

Together, the use of diatom indices and aquatic macroinvertebrate indices accurately reflects the water quality of South Africa rivers and other aquatic ecosystems. However South African citizens have little awareness of the techniques which mostly used in a scientific or commercial monitoring context. With the correct adjustment to these techniques through digital innovation and simplification of use, greater knowledge dissemination regarding water quality issues can be achieved and allow the public, once

informed, to stand as advocates for correct water management in the greater context of South Africa.

1.4) Challenges in biomonitoring in South Africa

Biomonitoring in South Africa provides the DWS with important information and data to incorporate into the REMP and similar monitoring frameworks to set RQOs to adhere and comply with the NWRS. However, logistical, environmental, socio-economic, political and technical challenges, most notably capacity constraints, taxonomic expertise, limited identification guides and weak infrastructure, hinder the application of biomonitoring aquatic systems (Graham *et al*., 2004; Bonada *et al*., 2006; Mangadze *et al*., 2019; Arimoro *et al*., 2024).

The landscape of South Africa is diverse, and many rivers originate from mountain ranges that are difficult to access and flow through privately owned land, which reduces the percentage of the river accessible to biomonitoring. The water quality of headwaters is important to determine and for use as a reference for downstream impacts, since pollution sources are diverse and can shape aquatic communities in many different ways (Alexander *et al*., 2007; Freeman *et al*., 2007). Site access downstream is often difficult with the infrastructure of South African roads prohibiting access. Additionally, the lack of taxonomic and field expertise, as well as limited funding and a lack in technical support further reduces the data generation through biomonitoring (Dalu & Froneman, 2016; Mangadze *et al*., 2019; Adom & Simatele, 2024). The lack of governance in the DWS and local municipalities can lead to failure in large-scale water monitoring programs and rehabilitation projects, while existing standards, as laid out in the NWA, are not adhered to and not enforced at local levels and provides a huge challenge for biomonitoring (Adom & Simatele, 2024). In addition to logistical and political challenges, socio-economic challenges unfortunately exist. There is a lack of equitable education and access to water, cultural barriers exist in the sense that education is not offered in traditional languages and contributes to community illiteracy and hinders community-based monitoring (Graham *et al*., 2024; Nqowana *et al*., 2024).

These challenges hinder the implementation of biomonitoring to determine the water quality of South African rivers which in turn hinder the implementation of management strategies to alleviate impacts from pollution.

1.5) Addressing technological and methodical challenges in diatom and aquatic macroinvertebrates indices

The diatom and macroinvertebrate-based indices to assess ecological conditions are powerful tools, however, some limitations exist that hinder the implementation and widespread use of such indices. Traditionally, macroinvertebrate indices are calculated manually, however with the development of the miniSASS application calculations have been done automatically, at least for miniSASS (Koen *et al.*, 2023; Pattinson *et al*., 2023). The calculation of SASS5 scores remains manual and if data is not stored electronically, this can potentially lead to data loss. Diatom indices are calculated using licensed software, OMNIDIA, on local devices and lack the dissemination of data and if these local devices are lost means the data generated and archived on these devices are also lost, because there is no central repository for diatom index scores, at least regionally. However, there are online repositories that have publicly accessible data on diatom index scores, especially in central Europe (Stenger-Kovács *et al*., 2023) which could be used as model for South Africa in the future.

With the implementation of technological tools such mobile applications and the advent of cloud-based storage, the collection and curation of data is improved both temporally and spatially (Koen *et al.*, 2023; Pattinson *et al*., 2023). Additionally, the generation of metadata such as GPS coordinates, site photos, site descriptions, collection dates, etc. can contribute to the robustness of data stored online since diatom samples are frequently found without metadata, especially date of collection. Without metadata, the use of data is limited and repeatability in experiments or temporal and spatial assessment is hampered (Szczepocka & Żelazna-Wieczorek, 2018). Therefore, when indices are calculated automatically, and results are deposited into cloud-storage without the risk of data loss together with describing metadata, the applications for diatom and macroinvertebrate indices in biomonitoring potentially improves. Additionally, taxonomic identification of diatoms and macroinvertebrates is a hinderance that can be overcome or aided by machine learning and image recognition to improve the taxonomic data available

for species. This will potentially improve the calculation of regional indices that allow for better standardization and comparison between regions and within regions. However, machine learning is still a novel tool used for species identification for diatoms and macroinvertebrates, with large datasets needed for training and validation (Pattinson *et al*., 2023; Szczepocka & Żelazna-Wieczorek, 2018). Additionally, data generated by mobile applications and standardized diatom software must be validated since data generated by citizen scientists can be unreliable and taxonomic identification of diatoms species lacks standardization, hindering the use of data generated (Baker *et al*., 2021). Therefore, by using digital tools such as mobile applications, computer software and machine learning models, data generation is improved and bolstered with the addition of metadata and the decreased risk of data loss through online cloud-based storage. The implementation of such tools will improve the efficiency of biomonitoring and therefore creation such tools for regional use is of utmost importance to bolster the use of such tools on a global scale.

1.6) Digital innovation in miniSASS and diatom indices for the purpose of knowledge dissemination

The digital divide in South Africa still hinders the full implementation of digital tools for knowledge dissemination. However, this divide is rapidly closing as more towns, schools, and communities gain internet access through widespread Wi-Fi networks and data coverage by mobile providers (Gillwald *et al*., 2018). Smartphones have also become increasingly common, the cost of smartphones has also decreased to the point where owning one is no longer a luxury but a standard for many people, however, only 44% of South Africans owned a smartphone in 2023. Therefore, although prices are decreasing, the affordability of smartphones still hinders the dissemination of information through digital tools. (NielsenIQ, 2018; GMSA, 2024). Regardless, the development and implementation of digital tools for biomonitoring remain feasible and timely.

The design and functionality of digital tools, especially mobile applications, must be streamlined, user-friendly, and efficient. Mobile applications should be capable of effectively conveying information on water quality if they are to serve as reliable tools for knowledge dissemination. It is essential that these applications include key features such as localized maps displaying water quality issues and survey sites, as well as information

pages containing training tools and relevant background material (Compas & Wade, 2018; Nqowana *et al*., 2024; Saleen *et al*., 2024). These resources can guide users before conducting a survey or simply inform them about water quality concerns in their area.

Clear explanations of survey techniques and instructions for conducting surveys correctly are highly important to ensure that the data generated is reliable and scientifically valid (Malthus *et al*., 2020). Effective identification tools and accurate index calculations are also crucial to minimise taxonomic misidentification and to improve the quality of resulting index scores. Communication between mobile applications and a central online repository is necessary for secure data storage and to prevent data loss (Kosmala *et al*., 2016; Balázs *et al*., 2021; Nqowana *et al*., 2024). Additionally, explanatory information on the generated index scores should be provided to help users interpret water quality results and understand what actions can be taken to improve conditions.

Digital tools for diatom index calculations are available but remain largely limited to a single licensed software package (OMNIDIA) in South Africa. While OMNIDIA offers robust calculations and efficient data management, it is outdated, having been developed in the 1990s (Lecointe *et al*., 1993), and is not particularly user-friendly, although later versions now exist that have corrected many of these issues. Users without access to the paid version are often required to send results to other researchers, which delays data processing and may compromise data integrity. Furthermore, the format required for data input is time-consuming and lacks modern innovation.

Developing a simple, freely available computer program capable of reading Excel-generated species matrices and automatically calculating diatom index scores, with explanatory graphs and tables, could potentially improve data throughput and reduce dependence on OMNIDIA for South Africans. Including ecological information about diatom species within such software could help explain index scores and guide informed decisions on pollution management. Moreover, incorporating indices specifically designed to reflect nutrient concentrations or salinity levels could further aid in identifying the nature of water quality problems, improving the management of aquatic ecosystems.

The development of free, easy-to-use digital tools, such as mobile applications and computer software, that can reliably generate index scores and securely store data can

play a major role in bridging the knowledge gap surrounding water quality issues in South Africa. These innovations can enhance public awareness, promote citizen science, and support the broader implementation of biomonitoring as a tool for sustainable water resource management.

1.7) Knowledge dissemination to the public of South Africa concerning water quality

Determining, maintaining and improving water quality of South African rivers and impoundments is of utmost importance if the country is to work toward fulfilling SDG6. Increasing pressure is being placed on water resources with a growing economy leading to increased agriculture, mining, urbanisation and pollution (GCIS, 2022). Public awareness of these issues is high in areas where the communities are faced directly with the impacts of these practices, however, the broader public awareness is lacking and effective dissemination of water quality information will empower citizens to participate in citizen science projects and become advocates for correct water management of water resources (Matta *et al*., 2020; Graham *et al*., 2024).

There are efforts in public knowledge dissemination on water quality with government programs such as WESSA, WaterCAN and Adopt-a-River having wide outreach and the involvement of NGOs and schools informing children on river health (Graham *et al*., 2016; Erasmus & Bega, 2025). However, mobile applications can help bridge this gap by introducing a digital, fun-to-use tool that can inform children and the public without the need to spend money on outreach programs and initiatives. Mobile applications can help improve the coverage of water quality issues and improve data generation of water quality assessment with the development of national water quality maps.

However as stated above, the digital divide is still relevant with many citizens not having access to smartphones or the internet, especially in rural areas where water quality issues or often worse (Wyrzykowski, 2023). Additionally, lack of data sharing between local governments and municipalities, and institutions hinders this knowledge dissemination (DWS, 2022d; Steyn, 2022). Technical jargon is also not directly translatable into scientific language and in addition, the availability of information in all official languages of South

Africa severely hinders the effective dissemination of water quality information (Brown *et al*., 2011; Nqowana *et al*., 2024).

Partnerships between schools, NGOs and local municipalities can help bridge this gap, with the addition of including translated information available in all 12 official languages. Social media can also provide a powerful initial dissemination of available mobile applications and outreach programs to get people involved in citizen science projects.

Therefore, with the introduction of social media and mobile applications, and the incorporation of translating services into these tools, the dissemination of water quality information could be improved, and citizens can be empowered to act as stewards of correct water quality management to help combat the increasing pressures placed on water resources.

1.8) Study Rationale

Water quality in South Africa is a particular concern considering the arid climate of the country, the lack in natural lakes and the reliance on rivers, man-made dams and wetlands as primary water sources. Rivers serve as the main conduits of water for industrial, domestic and agricultural use and monitoring the quality of rivers is a particular challenge for South Africa and is to sustain the economy. Wetlands are natural sinks and biodiversity hotspots in South Africa; however, they are increasingly under threat by agriculture, mining and construction (including dams). Biomonitoring of rivers and wetlands has been a valuable addition to chemical methods, however, survey techniques and index calculation remain a challenging aspect of biomonitoring, particularly using diatoms and aquatic macroinvertebrates. Diatom index calculation is done locally using licensed software (OMNNDIA v 5.3) whilst macroinvertebrate indices are physically calculated during field surveys and uses physical dichotomous keys for species identification. Software such as AQEM/STAR Ecological River Classification System (ASTERICS) and the River Invertebrate Prediction and Classification System (RIVPACS) for the calculation of macroinvertebrate indices and scores is available, however, these were developed for use in Europe and the UK and is not adapted for use in South Africa. The lack of digital biomonitoring tools in South Africa results in low data throughput, fragmented data storage, and frequent data loss. Digital tools designed for use in South

Africa can automate index calculation, improve data accuracy, and facilitate centralized data storage. Additionally, these tools enhance data accessibility, knowledge dissemination, and long-term data management. By implementing digital systems for index calculation and data handling, the efficiency, reliability, and scalability of river and wetland health assessments in South Africa can be significantly improved.

1.9) Problem statement

Aside from licensed software such as OMNIDIA v 5.3, there are no digital tools available for diatom and macroinvertebrate index calculation and data management. Digital tools are required for index calculation, data storage, and data management to improve biomonitoring efficiency and knowledge dissemination. New regional diatom indices for rivers and wetlands need to be formulated and incorporated into these digital tools.

1.10) Aim

To simplify and improve the efficiency and accuracy of biological indices used for river health monitoring by incorporating newly designed digital tools.

1.11) Objectives

1.11.1) Technological upgrade of MiniSASS

1. Development of a machine learning model for the automatic identification of macroinvertebrate groups used in miniSASS.
2. Development of a mobile application that incorporates automatic identification and data capture for macroinvertebrate surveys to improve the applicability and ease of use of MiniSASS as a citizen science tool.
3. Development of effective MiniSASS training tools that interface with the mobile application (training videos and interactive digital classification key, etc.).

1.11.2) Diatom Index software

1. Develop diatom indices using the unimodal distribution of species across environmental variables by inferring the optimum environmental condition and tolerance of taxa to ionic composition and nutrient concentrations (dissolved inorganic nitrogen and orthophosphate) using weighted averaging.

2. Develop diatom indices using the optimum environmental conditions and tolerances for taxa with nutrient, ionic composition and organic content inferred from expert knowledge.
3. Develop a diatom index for detecting AMD disturbance in wetlands.
4. Develop a new front-end input for diatom data capture, storage and index calculation to potentially replace the OMNIDIA platform which incorporates the newly created indices as well as the IPS.
5. Develop a simple executable program that calculates the wetland index for AMD disturbance.

CHAPTER 2: DIGITAL INNOVATION OF MINISASS AND THE INCORPORATION OF MACHINE LEARNING FOR AUTOMATIC IDENTIFICATION OF MACROINVERTEBRATES.

2.1) Introduction

2.1.1) The introduction of aquatic macroinvertebrate indices in South Africa

In South Africa, the introduction of macroinvertebrates as a standard biomonitoring tool within the scope of the River Health Programme, now known as the River Eco Status Monitoring Program (REMP) was during the early 2000s. Macroinvertebrates as a biomonitoring tool was formally introduced in 2004 through the development of SASS5 (Dickens & Graham, 2002).

The introduction of SASS5 was preceded by nearly three decades of work starting with the design of the first Biotic Index for river assessment in South Africa by Chutter (1972), this index was not often used as was extremely labour intensive (Dickens & Graham, 2002). In the 1990s, however, Chutter developed the first version of the South African Scoring System (SASS1) through adapting a method proposed by the Biological Monitoring Working Party (BMWP) as a tool for national pollution surveys in the UK (Hawkes, 1997; Dickens & Graham, 2002). SASS1 was focused primarily on macroinvertebrate families and was largely experimental. Through the development of SASS1, newer versions were developed and tested through the introduction of South African-specific sensitivity scores for macroinvertebrates in SASS2, refined taxonomic groups, consistent scoring and better sampling techniques in SASS3, and further refinement of taxonomic groups and sampling protocols in SASS4 (Chutter, 1994). SASS4 was upgraded to SASS5 by the inclusion of updated sensitivity scores for taxa, formal metrics such as a SASS score and an Average Score Per Taxon (ASPT) and standardized sampling protocols (Dickens & Graham, 2002). Furthermore, SASS5 was validated across South Africa with extensive application and was fully integrated into the River Health Program (RHP, now called the River Ecostatus Monitoring Program (REMP). SASS5 serves as the primary biomonitoring tool for riverine aquatic assessments and includes 91 families of aquatic macroinvertebrates. The efficacy of SASS5 is high and it continues to be the primary biomonitoring method in South Africa. SASS5 is a method

reliant on the identification expertise of the operator and requires training to effectively and reliably conduct a survey.

During the same time as the SASS5 development Graham *et al*. (2004) developed the mini Stream Assessment Scoring System (miniSASS) for the purpose of community engagement. miniSASS is a simplified version of SASS5 where 13 groups are included that encompass the 91 families used in SASS5. The focus was on engaging the community to take part in biomonitoring surveys using macroinvertebrates without the need for in-depth training and focusing on a rapid assessment method for early detection that could be followed by a full SASS5 survey if needed. This allows users to rapidly conduct surveys and engage in community projects and early pollution detection (Graham & Taylor, 2018). This method appeals to non-scientists and serves as an effective citizen science tool to help bridge the gap between knowledge dissemination from scientific research to public inquiry (Graham *et al*., 2004; Koen *et al*., 2023; Nqowana *et al*., 2024; Mickelsson *et al*., 2024).

2.1.2) miniSASS as a citizen science tool for rapid ecological assessment.

As mentioned, miniSASS is a citizen science tool that allows for rapid bioassessment of a desired river system in terms of water quality and pollution detection (Graham *et al*., 2004). The methodology of miniSASS requires users to sample a river at various locations (biotopes) for a certain amount of time to collect the aquatic macroinvertebrates present. The collected macroinvertebrates are then identified by using a dichotomous key. In the calculation of a SASS5 score, macroinvertebrates identified are classified into one of 91 families (Dickens & Graham, 2002), however, in miniSASS, the identified macroinvertebrates are classified into one of 13 overarching groups. Therefore, whilst simplifying the classification of invertebrates for non-scientists, the loss of taxonomic resolution can compromise the accuracy of scores generated. The score is calculated by using the assigned sensitivity scores in a simple average calculation on paper. The score is then interpreted and uploaded to the miniSASS website.

The simplification of a robust biomonitoring method such as SASS5 into miniSASS requires compromises in the sampling and identification of macroinvertebrates, the calculation of the final index score and the interpretation and archiving of these scores.

The duration of sampling is reduced in miniSASS from 2-3 minutes per biotope to a few minutes overall in whichever biotope is available (Graham *et al*., 2004). The sampling equipment such as waders is not provided and is often not available for citizen scientists and so the representation of macroinvertebrates in the system might not be as complete as when conducting SASS5 sampling. Furthermore, the calculation of a miniSASS score is done using groups and not families as in SASS5, thereby reducing the taxonomic resolution and compromising the accuracy of scores generated. Additionally, the calculation is reduced to a simple average calculation instead of using weighted averages (Graham *et al*., 2004).

These compromises in the methodology and calculations of miniSASS are further exacerbated through the lack of scientific training for citizen scientists. The identification abilities and ecological interpretation by citizen scientists may be somewhat lacking and can lead to the misinterpretation, and even miscalculation, of scores. Furthermore, the generated miniSASS scores are housed in the miniSASS website, and due to the mentioned limitations and compromises, these scores must be manually verified before release, and therefore, another bottleneck is created in the throughput of biomonitoring data generated using miniSASS.

These limitations can be alleviated by introducing a mobile application that contains all the guides, calculations and interpretations needed to conduct a miniSASS survey. Users can consult sampling guides, interpretations, descriptions of macroinvertebrate groups, ancillary information on why miniSASS is useful and what is does to increase their understanding of ecological monitoring and the importance thereof. The introduction of a digital identification key that replaces the dichotomous key, based on filtering groups using visible and easy identifiable features of macroinvertebrates can aid in the identification process.

Additionally, the application can house a machine learning model that aims to automatically identify the macroinvertebrates present (group level) to remove some of the human error present in identifying macroinvertebrates and to alleviate the bottleneck created by manual verification of results to automatically verify the miniSASS score on the miniSASS website (Koen *et al*, 2023; Pattinson *et al*., 2023).

2.1.3) The use of mobile applications for biomonitoring and surveying

With the introduction of smartphones, mobile devices became capable of capturing multiple forms of data such as coordinates, photographs, and timestamps, making them a feasible alternative to traditional paper-based data collection methods (Teacher *et al*., 2013; Lwin *et al*., 2014; Luna *et al*., 2018; Tabasi *et al*., 2024). Initially, mobile devices were mainly used for data capturing and recording observations rather than directly for biomonitoring. However, their use will lead to the collection of larger, more consistent datasets, which will help improve the quality and coverage of ecological surveys (Teacher *et al*., 2013).

The use of mobile applications in biomonitoring started with the migration of large online biodiversity repositories to mobile devices, which became possible with the development of more powerful smartphones. Some of the earliest examples include eBird (2009) and iNaturalist (2024), both of which started as web-based citizen science platforms before being adapted for mobile use (Di Cecco *et al*., 2021; Sullivan *et al*., 2009). These applications allowed users to collect and upload species observations directly from the field, forming the basis for modern mobile-based biomonitoring. The world largest online repository for biodiversity data, the Global Biodiversity Information Facility (GBFI), hold species occurrence records and more than 50% of these records are contributed through citizen science observations, like iNaturalist and eBird (López-Guillén *et al*., 2024; Truong & Van der Wal, 2024).

As mobile technology develops further, mobile applications are becoming real-time biomonitoring tools, capable of uploading and storing data on centralized repositories. This is allowing researchers and citizen scientists to generate more reliable data while improving accessibility and data sharing across different regions. Applications such as FreshWater Watch and GangaWatch are examples where standardized sampling protocols and built-in scoring systems are used to assess environmental conditions (Sandha *et al*., 2017; Bishop *et al*., 2025). The integration of these tools has transformed biomonitoring from traditional, localized surveys into a more connected and data-driven approach that supports large-scale environmental monitoring and management.

2.1.4) The application of Artificial Neural Networks (ANN) in ecology

Artificial Neural Networks (ANNs) are a type of machine learning (ML) that is inspired by biological neural networks (brains) (Borowiec *et al*., 2022; Qamar & Zardari, 2023). ANNs are part of artificial intelligence (AI) in the sense that it is a way to teach a machine or computer to complete certain tasks or make predictions. Essentially, an ANN learns a pattern in data given to it and makes a prediction as an output (Quetglas *et al*., 2011). In 1958, Rosenblatt invented a single layer perceptron, which is the original model for a single neuron and the simplest form of an ANN. This single layer perceptron takes one or many signal inputs and gives an output as a prediction (Rosenblatt, 1958). Minsky & Papert (1969) later evaluated perceptrons and highlighted limitations of single layer perceptrons. Rumelhart et al (1986) developed a backpropagation algorithm that allowed multi-layer perceptrons to be trained efficiently for non-linear problems, which gave rise to the wider application of ANNs during the 1990s (Quetglas *et al*., 2011). Broadly speaking, in the field of ecology, the use of ANNs were introduced by Lek & Guégan (1999) and Lek & Guégan (2000) who discussed ANNs as a tool in ecological modeling. Thereafter authors like Olden & Jackson (2002) adressed the applicaiton of ANNs in ecology by providing important citique and adresing barriers by making ANNs more interpretable and practical in ecology. The use of ANNs in ecology has increased, with researchers using different types of ANNs to complete various tasks and to make predictions about data structure depending on the research questions (Chen *et al*., 2020; Pichler & Hartig, 2023).

From the 1980s onwards, ecologists were faced with a world of increasing environmental problems, such as climate change, biodiversity loss, pollution, and overexploitation (Ehrlich & Mooney, 1983; Sala *et al*., 2000; Millennium Ecosystem Assessment, 2005; Dudgeon *et al*., 2006; Diaz *et al*., 2019; Nikolaou & Katsanevakis, 2023), and the need to analyse data and make future predictions was becoming more important since traditional statistical methods were too time-consuming and inaccurate to make such complex predictions (Quetglas *et al*., 2011). Using traditional statistical methods such as multivariate analysis in ecology was still challenging due to the nature of ecological data, which is highly complex, often non-linear, bulky, and full of noise. ANNs offered a viable alternative to traditional methods since they are powerful tools that can manage large,

complex datasets and make accurate predictions even if relationships between variables are non-linear or unknown (Quetglas *et al*., 2011).

A challenge faced by those using ANNs was, and still is, the availability of large datasets. To train an ANN, a large dataset is required, and obtaining such datasets was difficult. Additionally, additional datasets are needed to test the training of the ANN; however, researchers used cross-validation and jack-knifing to overcome this problem (Guisan & Zimmermann, 2000). There are many types of ANNs, of which the Multilayer Perceptron (MLP) is one of the most popular types used in the early 1990s and 2010s to make predictions on species presence and absence at sites or used as regression to illustrate correlation between data. For example, Feio *et al.* (2014) used a combination of ANNs to make predictions of species assemblages at monitoring sites to better understand the difference between reference and disturbed conditions, thereby improving the bioassessment of freshwaters. Therefore, ANNs have become valuable tools for handling the complexity and volume of ecological data that traditional analytical approaches struggle to manage efficiently.

Convolutional Neural Networks (CNNs) are a type of ANN specifically designed for image recognition. CNNs have been successfully applied to many types of organisms, including aquatic macroinvertebrates (Arje *et al*., 2020; Milošević *et al*., 2020; Høye *et al*., 2022). Høye et al. (2022) used CNNs to automatically identify aquatic macroinvertebrates based on images taken using specialized techniques such as the BIODISCOVER (Biological specimens Described, Identified, Sorted, Counted and Observed using Vision-Enabled Robotics) machine for imaging specimens and image analysis. These studies have been extremely successful and demonstrate the applicability of machine learning using ANNs, specifically CNNs, for image recognition in ecology (Ärje *et al*., 2020; Milošević *et al*., 2020; Høye *et al*., 2022).

Using CNNs, researchers are also able to generate large taxonomic datasets that are traditionally tedious and take long to compile. With high accuracy in identification and prediction of taxonomy, machine learning can generate large, accurate, and trustworthy datasets of organisms identified to species level, thereby increasing the throughput in

data generation that can be used for other purposes, including retraining neural networks (Norouzzadeh *et al*., 2018; Eichinski *et al*., 2022; Binta Islam *et al*., 2023).

Although using ANNs shows much promise in the field of ecology, covering a wide range of ecological research questions, the development of a machine learning solution is a difficult task, with large datasets required, computation time, and computing power being potential drawbacks (Christin *et al*., 2019). Therefore, ecologists should evaluate all aspects of machine learning in ecology before deciding to implement it. However, collaboration between ecologists and other disciplines, such as computer science, can help mitigate some of these constraints.

2.1.5) Image recognition using Convolutional Neural Networks (CNN's)

Machine learning algorithms are sets of rules or instructions used by an artificial intelligence to conduct certain tasks or processes. Deep learning is a subset of machine learning and is aimed at teaching a computer or machine to process data in a way that is inspired by the human brain and to performs tasks or carry out rules simulating the human brain (Sarker, 2021). To process data like humans, deep learning techniques use a neural network, which is a layered structure with interconnected nodes or neurons to recognise patterns in data. A neural network can typically be divided into types of networks to perform tasks pertaining to the nature of the input data (LeCun *et al*., 2015; Goodfellow *et al*., 2016). A CNN is typically used in image recognition and processing due to its capability of recognizing patterns in images and is applicable for images of aquatic macroinvertebrates (Joutsijoki *et al*., 2014; Rawat & Wang, 2017; Milosavljević *et al*., 2021; Serna López *et al*., 2020; Ärje *et al*., 2020).

A CNN functions by applying two layers to images being processed, a convolution layer and a pooling layer. A convolution layer is formed by images being scanned by use of a filter which extracts important features from the images given, these extracted feautes are then pooled into a next layer that contains important features from input images (LeCun *et al*., 2015; Rawat & Wang, 2017; Alzubaidi *et al*., 2021). The convolution is again applied to further extract important features from the next layer, these are again pooled and the process continues until a set of features are obtained used to accurately distinguish between image inputs. The process of convolution and pooling requires a

learning process to instruct the algorithm on which features are important and must be used to distinguish between images, additionally large datasets of images must be used for higher accuracy (LeCun *et al*., 2002; Goodfellow *et al*., 2016). This learning process can be either supervised or unsupervised. Supervised learning techniques require a previously defined set of features given to the algorithm by which they conduct the convolution and pooling process. Unsupervised learning on the other hand, hands over the process of feature extraction to the algorithm by which it automatically chooses important features of images to extract information from images (Bengio *et al*., 2013; Alzubaidi *et al*., 2021). Supervised and unsupervised learning includes a training process to ensure the automatic feature extraction is accurate. The training process includes three steps, training, testing and validation, further explained in section 2.2.3.2.

2.2) Methodology

2.2.1) Overview of the development process

The development process is split into three main parts, creating the machine learning (ML) model to automatically identify aquatic macroinvertebrates and classifying them, creating the digital classification key, and creating the android mobile application that will house the model and the key. The mobile application also has the full capacity to conduct a miniSASS survey without the use of the ML model. As mentioned before, the ML aspect included in the application has a primary purpose of alleviating the bottleneck effect created by manual verification of site scores. However, including the ML model introduces the concept of using CNNs to aid the throughput of data generated through biomonitoring surveys.

2.2.2) Mobile application

The development of the miniSASS mobile application was aimed at improving the process of conducting a miniSASS survey, to increase the accuracy of results generated by providing training tools, a digital key and ancillary information as well as to explore the idea of using machine learning to automatically identify macroinvertebrates.

All elements in the mobile application must behave synergistically to ensure the operation of the application is easy to use and easy to navigate. It is important to create a mobile application that is easier to use than the current technique, otherwise a simple surveying

technique becomes unnecessarily complex and difficult. During the present study the development of the application has progressed through the creation of a beta version and a final version based on the results from beta testing. After the final version was created, additional testing and field surveys were undertaken, to streamline the application, only the final version of the mobile application is presented in this dissertation. The mobile application was developed using android studios and the tools housed in the platform.

2.2.2.1) Architecture

The layout of the mobile application includes three main elements which are divided into three pages, one for each element. A map where users can view previous assessments, either from their own contribution or those of other users, a site assessment page where users can add a new site as well as a miniSASS survey, and an about page where all the necessary information on how to conduct a miniSASS survey (training tools) is located. Additional information is also given on how to practice safety when conducting a survey and the value of water quality monitoring. A glossary to explain relevant terms used during the assessment process is also included in the about page. This layout allows users to view assessments, create new sites and new assessments as well as consult all the necessary information on miniSASS, including training tools and explanatory information.

The main operational aspect of the application occurs within the site assessment page. The initial display of this page lists all the field sites added by the user as well as an option to add a new site. Each site listing can be accessed to add a new assessment to that filed site, multiple assessments can be added to one site over time. The option to add a new site to the list prompts users to enter the details for that site and to capture an accompanying site image as well as any water quality measurement they can obtain.

2.2.2.2) Map

The map included in the application is a world map. The map can be panned and zoomed to navigate to the desired location. The map displays information such as geographical features (towns and cities), natural features (rivers) and previously created miniSASS field sites.

2.2.2.3) Field site creation page

The field site creation page is the main operational part of the application. Once users navigate to the site creation page, they are prompted to add a site. They will be asked to

input all relevant information and meta data relating to the site. Users are required to input the site name, the river's name, the name of the person creating the site, the date of creation, the location of the site, the river type, a site description as well as any additional notes. The location is automatically obtained through the mobile device by allowing the location services to be accessed by the mobile application. The location also functions offline, therefore, it is not necessary to have an internet connection when creating a new site, however, when uploading the site information and survey results, an internet connection is required.

Once the site is created, a new assessment can be added to that site. The assessment prompts the user to conduct the standard process of sampling the aquatic macroinvertebrate community as well as to measure water quality parameters if the user can do so. Once the operator has finished sampling macroinvertebrates and measured the water quality, they are prompted to enter the data measured in the water quality tab. Once they entered this information, they are prompted to take a picture of each of the macroinvertebrates they sampled and to select the appropriate group from a dropdown menu into which the organism is classified, if users are having difficulty with the classification of organism, they have the option to consult the provided digital identification key.

It is only necessary to include one image of each group, however, for training purposes and thorough archiving and recording, users may take as many images of each macroinvertebrate group as they like even if it belongs to a group already included, the calculation is qualitatively calculated, and so additional images do not influence the final score. This process is repeated until every group is recorded and the assessment is then saved, this is also when the machine learning model will generate its own score in comparison to the user generated score; users have the option to upload the images taken to the miniSASS website to contribute to the collection of macroinvertebrate images in the dataset used to train the machine learning model. The application will return to the assessment page where the images for all macroinvertebrates are listed, together with their identification and the machine learning identification as well as the miniSASS score generated for that site.

2.2.2.4) About page

The final page included in the mobile application is an about page. This is where the important information about miniSASS is housed. Information on the importance of miniSASS, how to conduct a miniSASS survey, how to calculate scores, how to identify macroinvertebrates, how to sample macroinvertebrates and how to practice safe sampling protocols is included. Additionally, training tools such as videos that explain the sampling procedure and how to calculate and interpret scores are included and embedded in the mobile application and can therefore be accesses offline without an internet connection. A glossary on the important terms used in the application and in the digital key are included as well as how to use the mobile application. Additionally, links to the miniSASS website is also included where users can manually upload their surveys should they wish to do so.

2.2.3) Machine learning (ML) model

The ML model is created using Convoluted Neural Networks (CNNs) where images are given as input, these images are augmented to extract specific features, through supervised learning. The learning process is repeated over certain time periods, called epochs, until enough features are extracted to accurately identify organisms and classify them into one of the specified groups. Therefore, the process includes the creation of the ML model through constructing neural networks to extract features, acquiring sufficient images to use as input for training followed by testing and validation of the classification. Thereafter the model is built into the mobile application, the user has no input into the classification of organisms by the ML model and therefore the comparison of the user generated score is completely unbiased.

2.2.3.1) Data acquisition

To obtain images of macroinvertebrates useful for feature extraction and training of the ML model it was necessary to conduct sampling as there are no publicly accessible databases for macroinvertebrate images. Therefore, thirteen sites were extensively sampled for aquatic macroinvertebrates according to the standard miniSASS sampling protocol. Six sites in the North West (Figure 2) and seven sites in KwaZulu-Natal provinces in South Africa (Figure 3) were sampled during September 2022. Metadata collected in tandem with sampling included site name, coordinates, site description,

groups of macroinvertebrates collected, and biotopes sampled (VEG = vegetation; SOC = stones out of current; SIC = stones in current; GSM = gravel, sand & mud) (Table 1).

Figure 2: Map showing the sampling locations for the North West province South Africa. The blue line indicates the flow of the Mooi River, white dots indicate sampling points along the river flow. A black dot indicates the location of the North-West University.

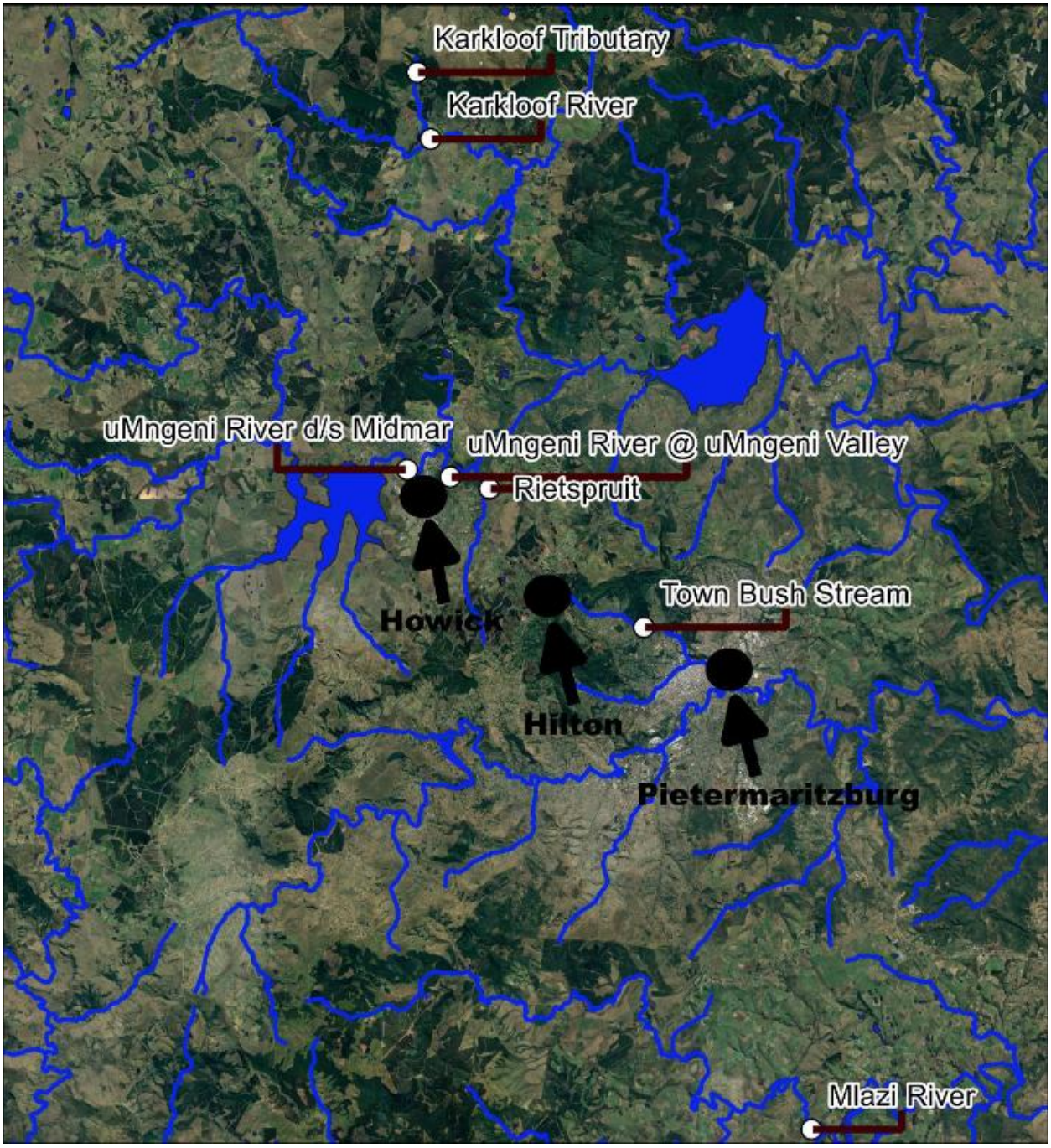


Figure 3: Map indicating sampling locations in KwaZulu-Natal province South Africa. Blue lines indicate rivers and dams, white dots indicate sampling locations, black dots indicate major town in the area.

Table 1: Sampling points for images of aquatic macroinvertebrates collected according to the standard miniSASS sampling protocol. Meta data captured include: site name, coordinates, site description, groups of macroinvertebrates collected and biotopes sampled (VEG = vegetation; SOC = stones out of current; SIC = stones in current; GSM = gravel, sand & mud).

Site name	Location	Site description	Macroinvertebrate groups collected	Biotopes sampled
S1 Mooi River	-26.68053, 27.0986	Medium depth, wide channel present with abundant riparian vegetation. Water quality moderate.	All groups - No stoneflies. No clams.	VEG, SOC, GSM
S2 Mooi River	-26.68119, 27.09838	Deep, wide channel with soft sediment. Many reeds present. Water quality moderate.	All groups - No stoneflies; No clams.	VEG, GSM
S3 Mooi River	-26.68431, 27.09954	Small, deep channel, slow flowing. Abundant riparian vegetation. Water quality moderate with some pollution.	All groups - Few dragonflies; No stoneflies.	VEG, SOC, GSM
S4 Mooi River	-26.68558, 27.10148	Small, shallow stream with many boulders. High stream flow. Little to no riparian vegetation.	Mayflies; Bugs & Beetles; Some Dragonflies.	SIC
S5 Mooi River	-26.68676, 27.10341	Heavily shaded stream, small channel width and depth. Little riparian vegetation. Some pollution.	All Groups - No Stoneflies; Many clams.	GSM
S6 Mooi River	-26.6935, 27.10305	Site within NWU. Many freshwater red algae. Strong current with many boulders. Some pollution.	All groups - No stoneflies; No clams.	VEG, SIC, GSM
Karkloof Tributary	-29.30106, 30.22779	Pristine mountain stream, shading abundant. Shallow stream with scattered shallow pools. Some riparian vegetation.	All groups (mostly stone-flies).	VEG, SIC, GSM
Karkloof River at Spitskop	-29.3336, 30.2353	Slightly impacted site, located under bridge. Abundant riparian vegetation. Small stream with sloped bank.	All groups.	VEG, SIC, GSM
uMgeni River downstream of Midmar	-29.49484, 30.22218	Large, wide channel. Mostly large boulders present. Little GSM. Some Riparian vegetation.	All groups - few stoneflies.	VEG, SIC, GSM
Rietspruit	-29.50413, 30.26829	Small to medium stream, small waterfall present. High biotope diversity. Slightly impacted site.	All groups - no stoneflies.	VEG, SIC, SOC, GSM
uMgeni River at uMgeni Valley Nature Reserve	-29.49844, 30.24635	Large wide channel, deep with strong flow. Many large boulders with abundant of algae.	All groups - no stoneflies.	VEG, SIC, SOC, GSM

Town Bush Stream	-29.5717, 30.35493	Many large boulders with abundance of algae. Little vegetation, abundant GSM.	All groups - Many crabs; No Stoneflies.	VEG, SIC, GSM
Mlazi River	-29.81584, 30.44901	Site near agricultural practice, small to medium stream. Biotopes abundant, little boulders.	All groups - Many stoneflies & Shrimp.	VEG, SIC, SOC, GSM

During the study, a total of thirteen thousand images of as many different aquatic macroinvertebrates were collected using a Nikon 2MP camera and Samsung galaxy mobile device, images taken include high resolution images and slightly blurred images to account for the possible user error in the acquisition of images during the use of the mobile application. Groups sampled included: 1) flat worms, 2) worms, 3) leeches, 4) crabs and shrimps, 5) stoneflies, 6) minnow mayflies, 7) other mayflies, 8) damselflies, 9) dragonflies, 10) bugs and beetles, 11) caddisflies (cased and uncased), 12) true flies, and 13) snails (Figure 4). Sampling of sites was thoroughly conducted to obtain as many individuals as possible as groups contain high morphological diversity with many families included.

Figure 4: Examples of collected images. (**A** – stoneflies; **B** – uncased caddisflies; **C** – cased caddisflies; **D** – snails; **E** – clams; **F** – shrimps; **G** – crabs; **H** – worms; **I** – trueflies; **J** – leeches; **K** – beetle larvae; **L** – damselflies; **M** – dragonflies; - **N** – minnow mayflies; **O** – other mayflies; **P** – flatworms; **Q** & **R** – trueflies; **S** – beettle larvae; **T** – beetles).

A minimum of 1000 images for each group were required for image processing and classification. Images of macroinvertebrates were taken in different orientations of the organism to capture all morphological features of the individual. Images were then pre-

processed by means of brightness adjustment to alleviate high contrast images and to highlight important features of macroinvertebrates (Figure 5). Additionally, manual cropping was done to isolate individuals and reduce noise.

Figure 5: Example of the adjustment and cropping process to prepare images for training. (1 - unedited image; 2 - brightness and contrast adjustment; 3 - cropping).

2.2.3.2) Development of the machine learning model

Machine learning algorithms are sets of rules that a machine learning model uses to accomplish a specific task. In this case the model must accomplish the task of classifying image of aquatic macroinvertebrates into one of the thirteen desired groups. Therefore, a model must be trained by providing images for each of the macroinvertebrate groups. Once it's finished training (learning) what these images are, another set of images are given to validate what the model has learned, after which parameters are adjusted within the model if the training was not as desired. After training and validation, the model was presented with a new set of images to test its learning process. The testing is the most important step of the process as it gives the true accuracy of the model in a real-world situation. The dataset of 13 000 images was split into training, validation and testing accordingly: 70% for training, 20% for validation and 10% for testing.

Machine learning models usually learn by means of a convolutional neural network (CNN), which is a network of nodes that function like the human brain, when learning new tasks. However, creating a new model with a CNN developed from scratch is not necessary. Base models exist that are already trained to accomplish specific tasks. In this case, a base model that is trained on image recognition is perfect, its knowledge on images (shapes, edges, colour, contrast, etc.) can be used as it has already learned and then applied to the images collected for the present study with some adjustments. The base model used is the VGG19, this model is trained on more than 1 million images from the ImageNet databate, this database contains a wide variey of image categories. This

knowledge can be used and essentially transfered to the scenario in question. Therefore, a base model (Figure 6) is utilised, and its knowledge is transferred (transfer learning) to a new head model specifically designed to recognise macroinvertebrates.

However, every aspect of the base model was not needed. The base model consists of pretrained layers and top layers (Figure 7). Essentially the pretrained layers are the features the model has already extracted from images, such as shape, edges, colour, contrast, etc. and the top layers are the final classification layers where the features extracted in the pretrained layers are applied to make a final prediction on an image, for example a dog or a cat. Consequently, the top layers are not needed, only the pretrained layers, as these features are the same features used to classify images in the broader sense, all images have a shape, edges, colours, contrast, etc. Therefore, the top layers are removed and only the pretrained layers are transferred to the new model. However, the pretrained layers are also frozen. Essentially this locks in the specific weights of the features it already knows, as it is not required that the model to relearn it's pretrained layers, so they are frozen or lockedwhen transferring them to the new head model.

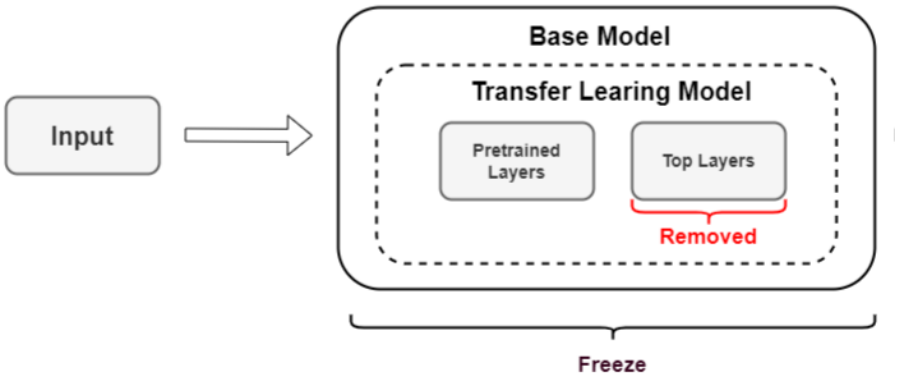


Figure 6: Illustration of the base model used for transfer learning.

Therefore, input of macroinvertebrate images is provided to the model, it already knows which features to extract through transferring the knowledge from the base model. The trainable section of the model is the head model, this part sits above the base model and utilises its knowledge to extract the features from the image used as input (Figure 7).

The head model is where 70% of the 13 000 images are provided as an input. Intially images of one specific group are given following on to the next groups sequentially.

Seventy percent of the images for each group (700) is used to train the head model. The knowledge of the base model is used to extract the most important features of the macroinvertebrate images (legs, antennae, eyes, wings, shape, etc) and pooled together. Imagine an image is split into different quadrants and within each quadrant there lies important features, max pooling is where only the most important features in each quadrant are used, pooling them into a new smaller image that contains all the most important features without losing information. This reduces the number of features used to train the model and makes the training process faster without compromising accuracy. The grid produced by max pooling is still a two-dimensional feature (a matrix), a CNN requires a one-dimensional input for training and therefore the 2D grid must be flattened into a single line for input. Therefore, max pooling and flattening creates a viable input of features for learning.

Whilst the model is learning which features are important for which type of macroinvertebrate, it memorises which features belong to which group, however, this memory can became saturated and no new features form part of the memory, and so important features are omitted and can lower the final learning curve of the model, and essentially this line flattens where no new features form part of the learning process. Therefore, the model is instructed to drop some of the features it has learned so as to force it to learn new features and not to use memorised features. The dropout rate is set at 0.5, therefore when learning new features, it randomly forgets 50% of the features it has previously learned, therefore, the pool of features it does learn only contains the most important features of all the features it can possibly use. These features are connected into a network, essentially a network consisting of a combination of features to predict the classification of a macroinvertebrate, this network is 512 layers dense, which means there are 512 neurons or nodes that can be used for high-level feature combinations, each node being a specific feature. The next network consists of 13 layers, which are the final classification groups, therefore, the combination of 512 features is condensed into one of the 13 layers, essentially taking a combination of the 512 features and making a prediction into one of the thirteen groups. Training the machine learning model is repeated over time periods called epochs. An epoch is the entire duration it takes for the model to see the training dataset once and make predictions based on what is has learned. It is necessary

to repeat the training process to ensure the accuracy of the final predictions improve. Initially, the training begins with random weights and as training over epochs progress, these weights are adjusted to minimise the error or loss between training and validation accuracy. Eventually, the predictions reach it's final accuracy and the training is halted to avoid noise fitting in the training data, essentially the training is halted to prohibit the model from learning new features based on noise which will decrease the accuracy of the model. Satisfactory training of a machine learning model has a tradeoff between training and validation accuracy and training and validation loss. As training accuracy improves the validation accuracy also improves, however, if the validation accuracy stagnates but training accuracy improves, the model is overfitting features and essentially memorising features instead of learning patterns, therefore the weights must be adjusted to reduce loss. Eventually training and validation accuracy plateaus and the training is halted.

Once the head model is trained, the combination of features that the model uses to classify an image of a macroinvertebrate into one of the thirteen groups is output into the feature stack which is uses in the final prediction as it is housed in the mobile application.

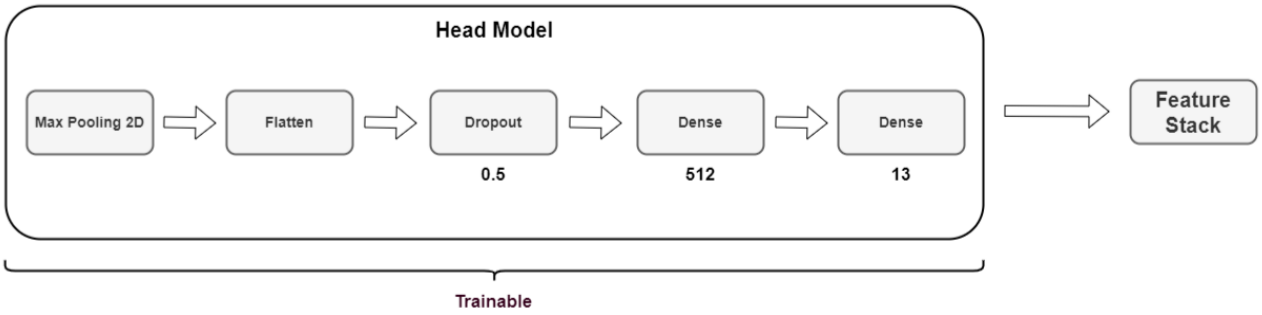


Figure 7: Workings of the trainable head model.

Therefore, the macroinvertebrate dataset containing a thousand images of each of the thirteen groups (13 000 images) was used to train a machine learning (ML) model to extract important features needed to accurately identify macroinvertebrates to group level. The model uses these features to train its image recognition of macroinvertebrates to classify them into one of the thirteen groups. The classification results are validated against known images of macroinvertebrates and then used to identify new unknown images given to achieve a final classification accuracy. The initial classification accuracy was expected to be between 70 and 80% because the number of images given are well

below the desired number of images used to train other machine learning models on image recognition, usually millions of images. However, with the accumulation of images collected during future testing, the model must be retrained in the future at 100 000 images to achieve a higher accuracy and then again at one million photos if possible.

2.2.3.3) The position of the machine learning (ML) model in the mobile application:

The ML model is included in the assessment stage of the site creation page. During the assessment, the user is required to take a photo of each of the macroinvertebrate groups they sampled. Should a user take multiple photos of the same group, only the presence of the group is used to calculate the score and not how many of the group were found. The scores produced from the classification of the macroinvertebrates are based on the selection by the operator as well as the selection by the model. Therefore, two survey scores are produced for comparison, one from the machine learning model and one from the user selection. This comparison serves a primary purpose. If the scores produced by the model and by the user are similar within a certain margin of error, the manual verification of the score is no longer necessary and the bottleneck is averted when adding new assessments to the website, however this can only be reliably used if the machine learning model is accurate in its classification. This accuracy depends on the number of images used to train the model, the more images that are used, the more accurate the classification becomes. The acquisition of the appropriate number of images is difficult to produce over a short period of time, and so the initial accuracy of the model will be less than desired. However, the initial accuracy of the model is high enough for its implementation into the application.

2.2.4) Digital classification key

The classification key currently used in miniSASS is a dichotomous key where users can navigate their way through the dichotomous tree by looking at the presence or absence of different morphological features (Figure 8).

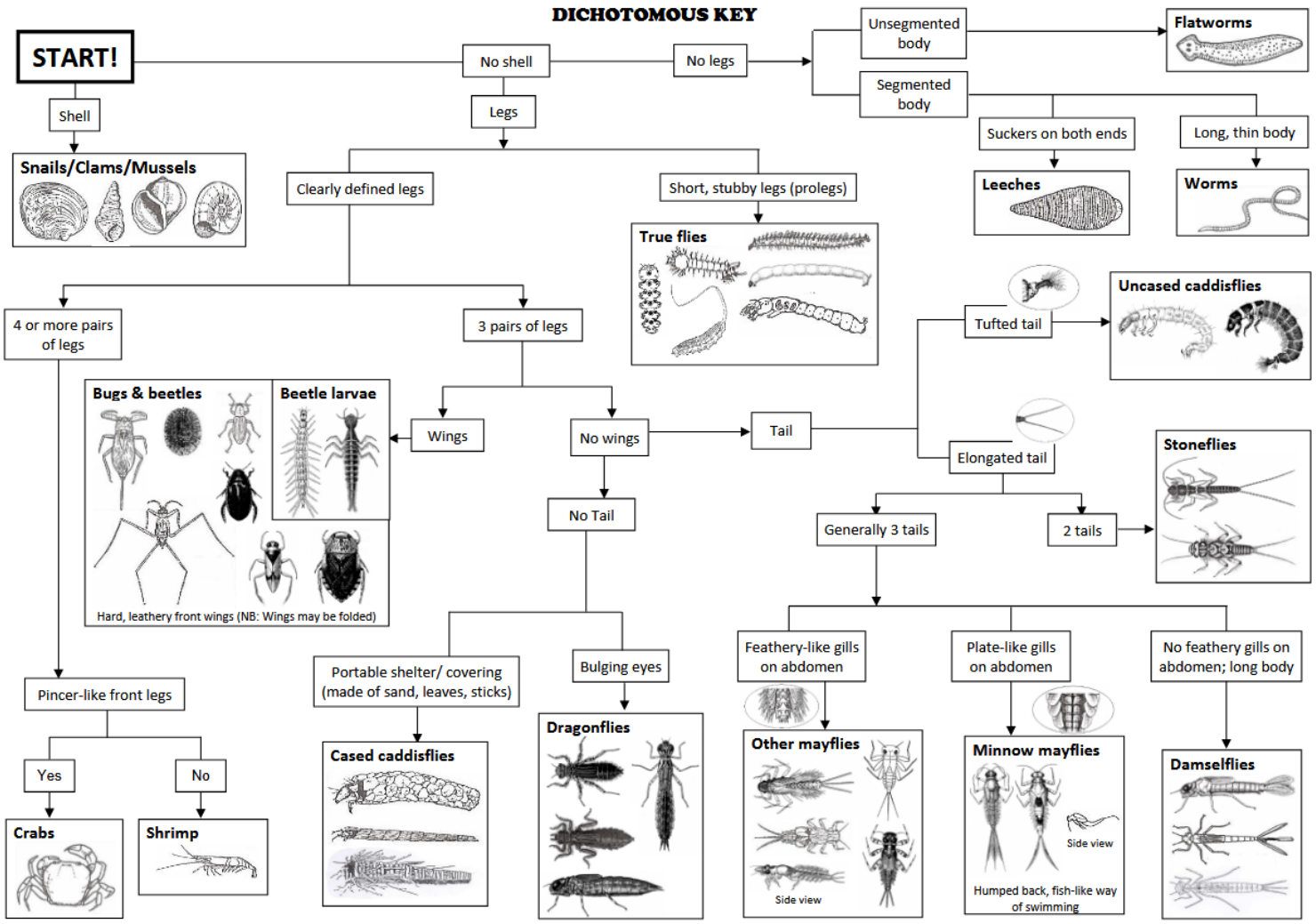


Figure 8: Example of the dichotomous key currently used in miniSASS surveys (minisass.org).

This key (Figure 8) prompts users to start at one predefined point and work through the key from there. Using this key, users are forced to start at the same point for each new identification and cannot start at any one of the criteria given. This creates an issue where some of these features are difficult to identify in some groups and can lead the user down the incorrect identification path when these features are not clearly visible or misinterpreted. For example, the pathway starts with the identification of a shell, which is relatively easy, thereafter if a macroinvertebrate has no shell the users are asked if they have legs, this is further split into prolegs and clearly defined legs. However, the prolegs listed as a feature are difficult to identify since they are extremely small and often overlooked and therefore this can create an identification issue early in the use of this dichotomous key. This can be alleviated through implementing a filter-based key by providing a list of features where users can select any one of these features, not limited to following a predefined pathway. Therefore, users are not prompted to look for something specific but rather select what they can identify and follow this method until they reach a classification.

The newly proposed key functions through a filtering system where each of the macroinvertebrate groups are assigned a combination of specific identifiers that are unique to that specific group (Table 2). The user is prompted to observe any feature they can identify in the organism they sampled and to select the appropriate feature from the list. The selection of a specific feature or identifier is accompanied by an image of that feature and will alter the list of groups accordingly to only include those groups that possess the selected feature. Other groups that do not have such features are excluded from the list and are unavailable for selection. The user then selects all the features they can observe in a specific organism and in so doing the list of potential groups becomes smaller until only one group remains for selection. Additionally, the users can deselect any of the features at any time if they are unsure, and groups that share the remaining features will be displayed. By incorporation such a filtering system, the complexity of a dichotomous key is avoided, and users can more easily classify organisms within the appropriate group.

Table 2: Unique identifiers for all thirteen macroinvertebrate groups used in miniSASS.

Group / Identifier	Bugs and beetles	Caddisflies	Crabs and shrimps	Damselflies	Dragonflies	Flatworms	Leeches	Minnow mayflies	Other mayflies	Snails, clams and mussels	Stoneflies	Trueflies	Worms
Shell										x			
Shelter		x											
Clearly defined legs	x	x	x	x	x			x	x		x		
Segmented body							x					x	x
Long, thin body	x	x		x								x	x
Appendages	x	x										x	
Three pairs of legs	x	x		x	x			x	x		x		
Four or more pairs of legs			x										
Elongated tail								x	x		x		
Tufted tail		x											
Short tail	x	x											
Plate-like gills								x					
Feather-like gills	x	x							x		x		
Leaf-like gills				x									
Bulging eyes					x								
Stocky body	x				x								
Antennae	x		x	x	x			x	x		x		
Suckers at both ends							x						
Wing buds				x	x			x					
Flattened body						x							
Short, stubby legs												x	
Rounded body	x												

From Table 2 a minnow mayfly can be classified using the characteristic of clearly defined legs, three pairs of legs, an elongated tail, plate-like gills, antennae and wing buds. A true fly can be classified by a segmented body, a long thin body, appendages and short stubby legs. Any of these features selected will present the macroinvertebrate group that shares these features. Therefore, each groups contains a unique combination of features that leads to their classification.

2.3) Results

2.3.1) Mobile application

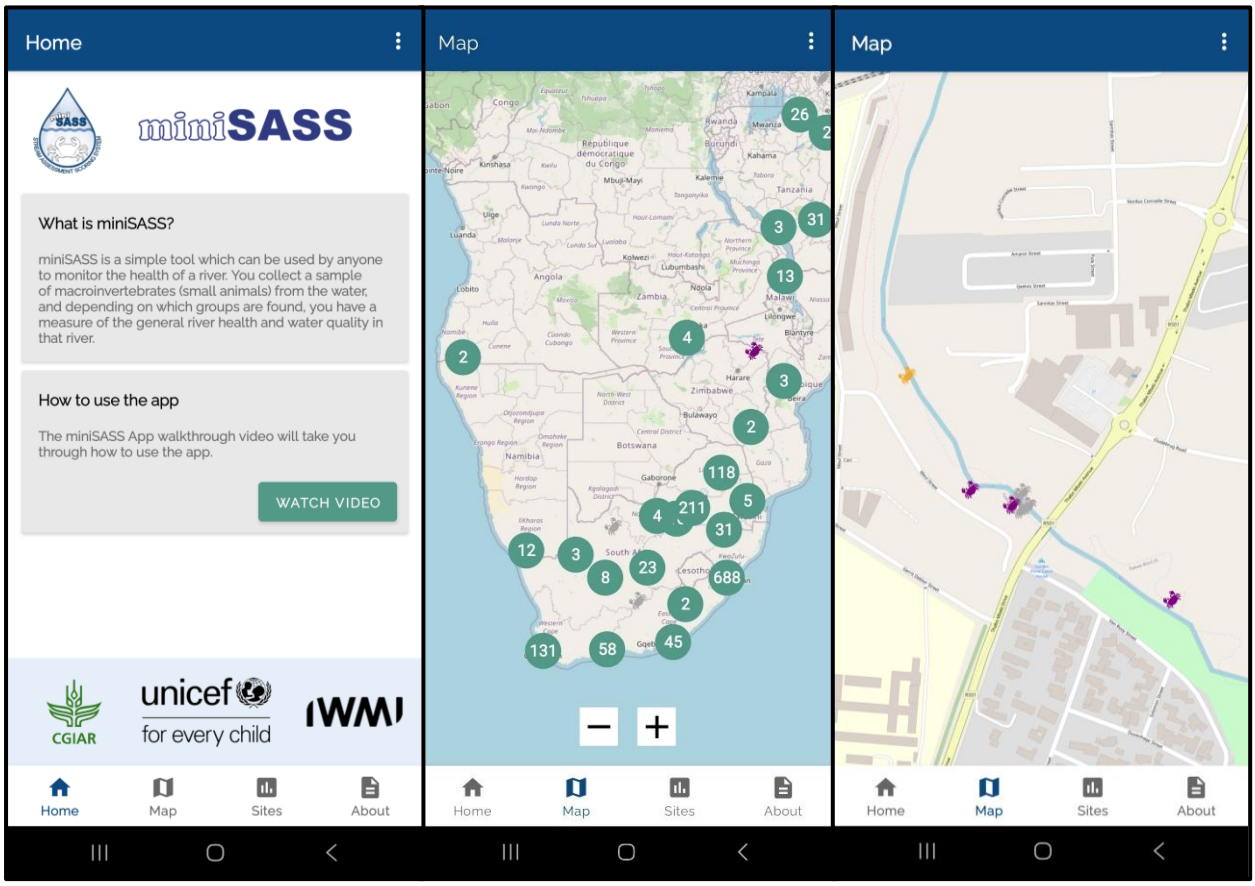


Figure 9: Initial landing page of the miniSASS mobile application (left), the map with sampling sites (middle) and locations of sampling sites with ecological state as indicate don the miniSASS website (right).

The miniSASS mobile application was developed in the present study to display four main pages that contain all the necessary information on miniSASS including the capability of conducting a survey. After launching the application, users are required to create a profile on the miniSASS website that links to their application. Any surveys conducted by the user will be linked to his/her profile and be displayed on the miniSASS website. After login, the user lands on the home page (Figure 9, left), this

page contains information on what miniSASS is and also provides a link to a video that explains how to use the application. Additionally, the sponsors that funded the development of the application are displayed. From the home page users can navigate to the map page, site page and about page. The map page displays a default google maps layout of the entire globe. Users can zoom into their desired locations where circles are displayed with a number inside, this number indicates the number of sites located in this general area, this ensures sites only load when the user zooms in far enough and not displayed as the map is opened, ensuring better functionality and faster loading. Once the user zooms in, the default crab icon is displayed as used on the miniSASS website.

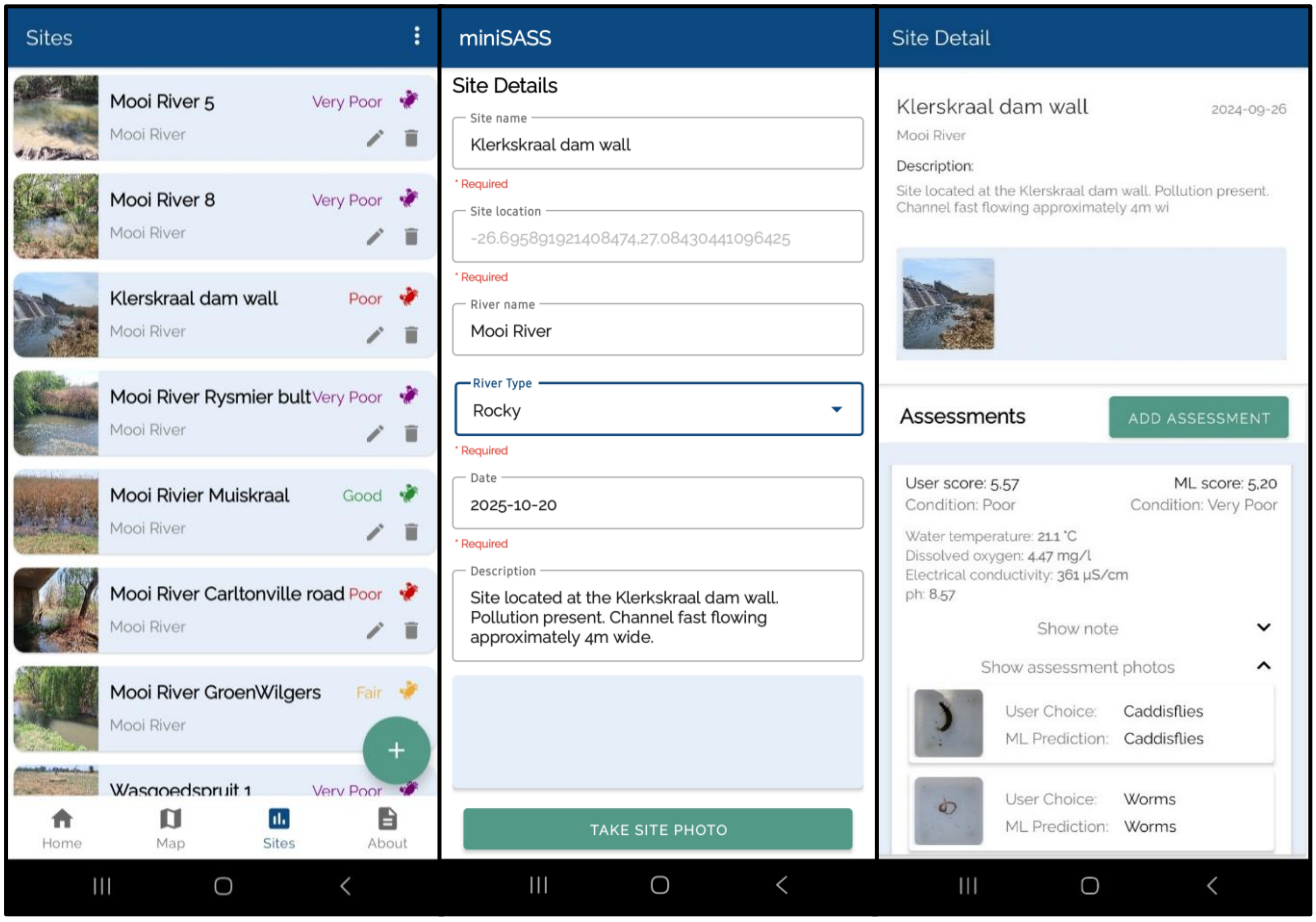


Figure 10: The field sites landing page (left), the fields sites creation page (middle) and the fields sites details page (right).

The field sites page is the main functional part of the mobile application. Here users can create sites and add site assessments. The sites page contains a list of all the sites created by the user with the corresponding scores for the latest site assessment. The site name is also displayed, together with the river name. Users have the option

to edit a site and/or to delete a site. Once a site is created it is uploaded to the miniSASS website. Users can select the (+) icon to add a new site whereafter they will be prompted to enter the site details as illustrated in Figure 10 (middle), the site location is automatically obtained through allowing the location services on the mobile device, the location is obtained without the need for an internet connection. Once the site is added, the users create a new assessment where they enter the water quality measurements Figure 10 (right) and proceed to capture images of the macroinvertebrates. Once they have captured all representatives of all groups sampled, the assessment is saved and a user and ML score is generated. The score is accompanied by images of the macroinvertebrates taken at the site as well as the user and ML prediction. Additionally, the water quality measurements are also listed (optional).

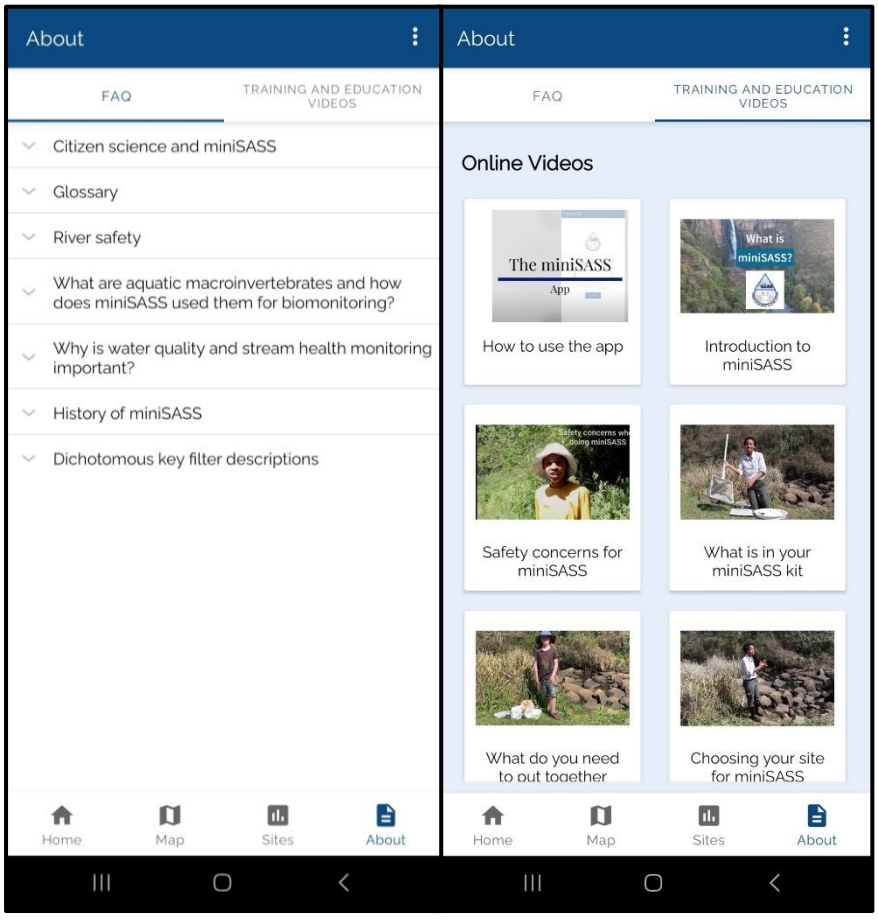


Figure 11: The about page contains the frequently asked questions and answers (FAQ) (left) as well as the training and education videos (right).

If at any point the users need information about miniSASS to help them complete a survey or to consult training videos they can navigate to the about page. Here a list of

frequently asked questions is given with the answers to those questions. Users can also watch supplementary videos that are included in the application and does not need an internet connection to play.

2.3.2) Machine learning model

The machine learning model was trained using 13 000 images of macroinvertebrate groups. 70% of images were used for training, 20% used for validation and 10% for testing. The training was run over 19 epochs (Figure 12) to improve training and validation accuracy and reducing training and validation loss. The initial training accuracy obtained was 69.27% whilst validation accuracy was 44.96%. This is well below the desired accuracy and so training was repeated until a desired training accuracy of 95.54% and a validation accuracy of 94.35% was obtained. From Figure 12 it is evident that the training and validation accuracy progression curves increase gradually and neither the training or validation accuracy plateaus before the end of the end of the 17th epoch which indicates the model has successfully learned from the training dataset and extracted features for accurate identification of macroinvertebrates. No further noticeable improvement was obtained from the 17th epoch, and training was halted after the end of the 19th epoch. Similarly, training and validation loss decreased from 6.98% and 2.38% to 0.13% and 0.14% respectively indicating a successful adjustment of weights during the learning process. Following training, the testing dataset was presented to the machine learning model and a final classification accuracy of 74% was obtained.

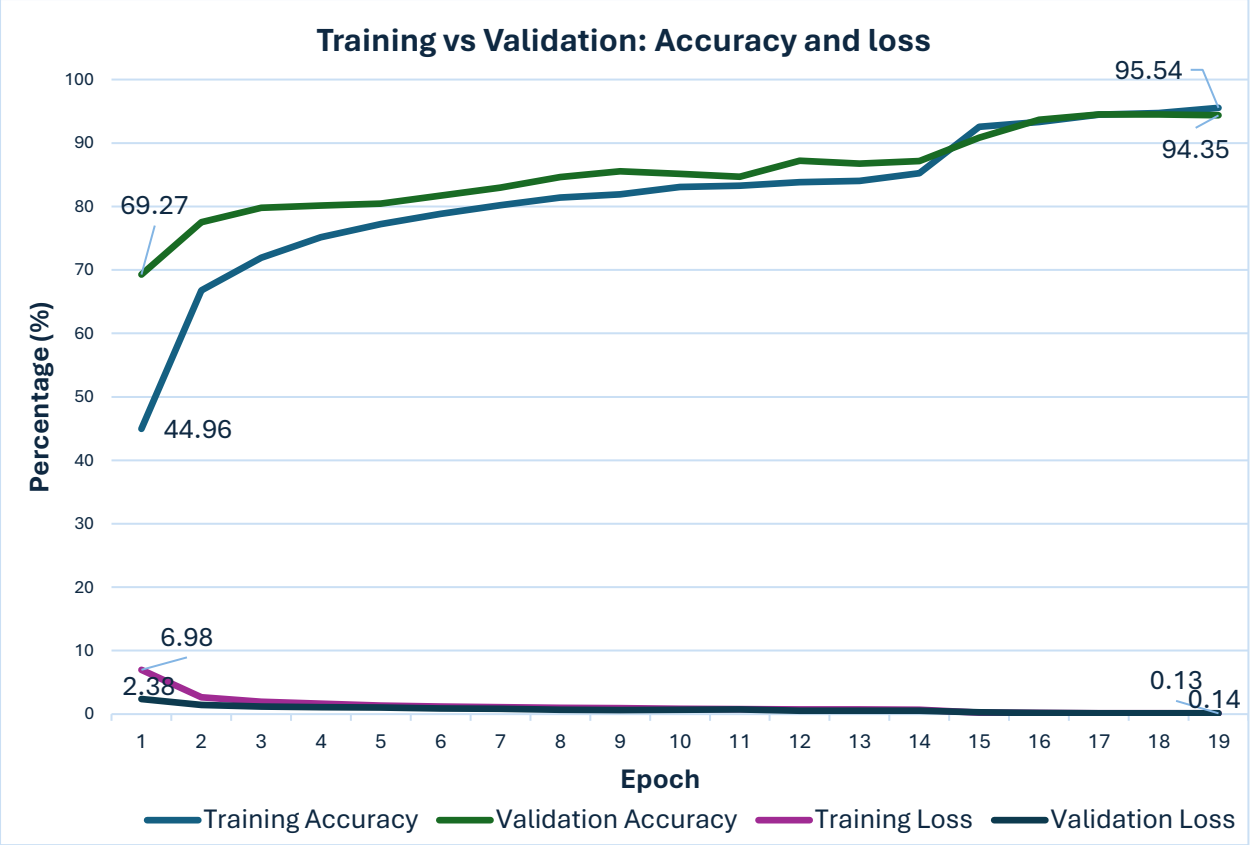


Figure 12: Machine learning model training and validation progress.

The final accuracy of 74% is below the desired theoretical accuracy of 95 – 99%, however, given the low number of images used to train the model, a final accuracy of 74% shows the model was successful in classifying macroinvertebrates however it can be further improved using larger numbers of images.

Whilst the training was successful, it is important to evaluate the performance of the model with regards to its failures and confusion in classifying macroinvertebrate groups. To better understand this a confusion matrix is used to illustrate the error in classification between groups (Figure 13). This graph is used to illustrate the error between the true given label and the predicted label from training. Each coloumn represents the predicted label and rows represent the true label. The confusion between the predicted label and true label is represented by a percentage or proportion between 0 and 1. From Figure 13 three clusters are evident, the first cluster in the top left corner of the graph illustrates the confusion of the model in classifying and distinguishing between bugs and beetles, caddisflies and crabs and shrimps. The prediction of bugs and beetles from images was accurate 50% of the time whilst the other 50% was split into 38% caddisflies and 12% other mayflies. Caddisflies were

accurately predicted 83% of the time whilst the other 17% was confused with true flies. Crabs and shrimp were accurately predicted only 40% of the time, with minnow mayflies also being predicted 40% of the time and bugs and beetles being predicted 20% of the time. The second cluster is formed on dragonflies and damselflies. Damselflies are accurately predicted 83% of the time with the other 17% being predicted as dragonflies. Dragonflies are only accurately predicted 50% of the time with damselflies being predicted 25% of the time and bugs and beetles the remaining 25%. The last cluster is on true flies. This group has the highest confusion in prediction of all groups. True flies are only accurately predicted 43% of the time, 29% of the time, leeches are predicted and 14% of the time flatworms and snails/clams/mussels are predicted. The remainder of the 13 groups are accurately predicted 100% of the time.

A possible explanation for this confusion in accuracy is the overlap in morphological features between groups as well as the high morphological diversity withing groups, such as bugs and beetles.

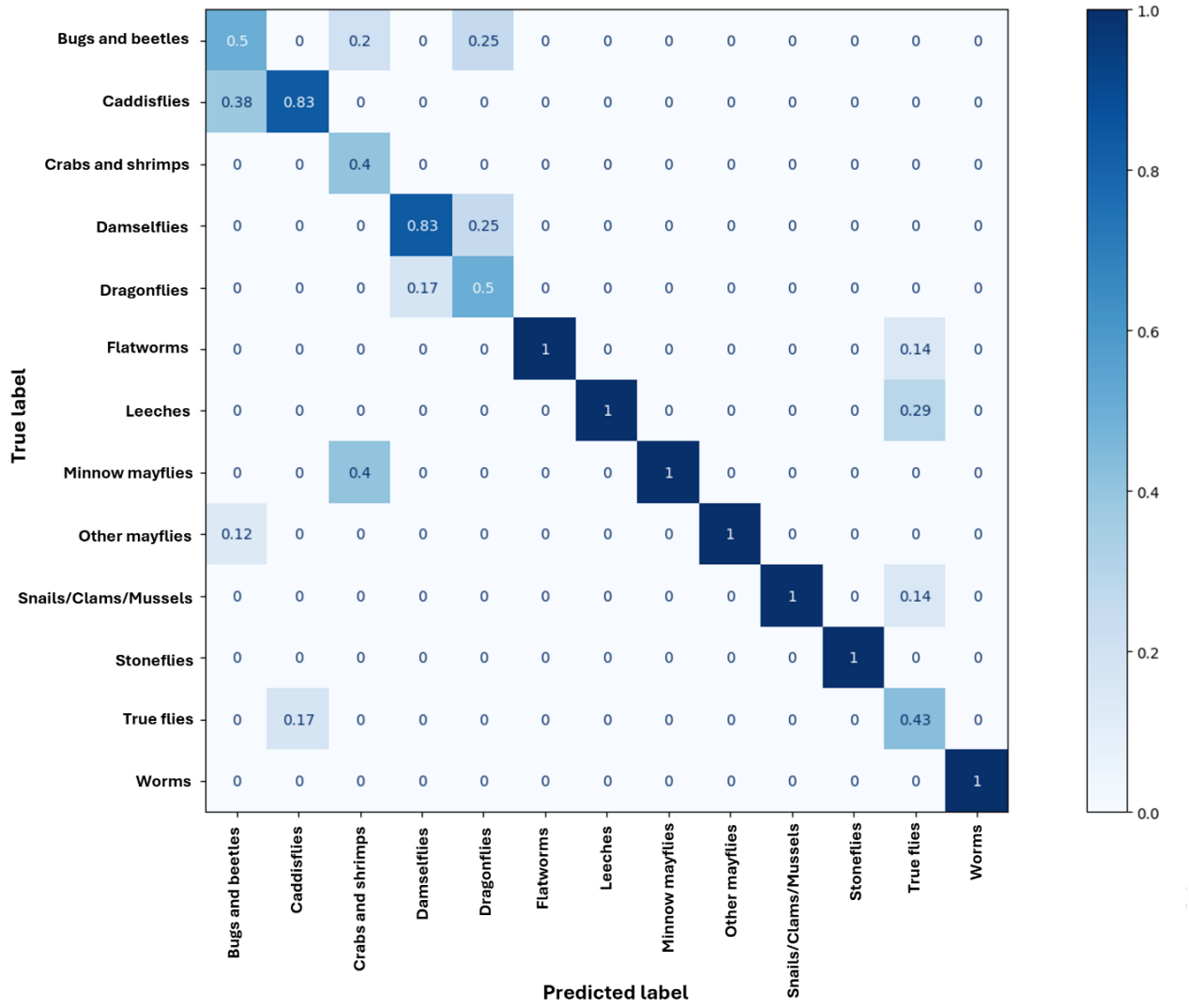


Figure 13: Confusion matrix for the trained machine learning model.

2.3.3) Digital classification key

During the assessment users are prompted to capture images of macroinvertebrates. After cropping and adjustment, they are required to classify the organisms into one of the thirteen groups by either selecting the classification from a dropdown menu, or consulting the digital classification key (Figure 14). The key contains a list of all features listed in Table 2 (Figure 14 left and middle) with a selection box accompanied by images that explain the relative feature (Figure 14 right). These identifying features are also listed in the FAQ section of the about page, under dichotomous key features, for users to consult before they begin with the assessment. It is necessary for users to familiarise themselves with all the ifentifying features before they conduct a survey. For clarification, some features are renamed to better explain the feature users are

searching for, example prolegs are renamed to short stubby legs. This provides a better explanation of the feature when users are searching for it.

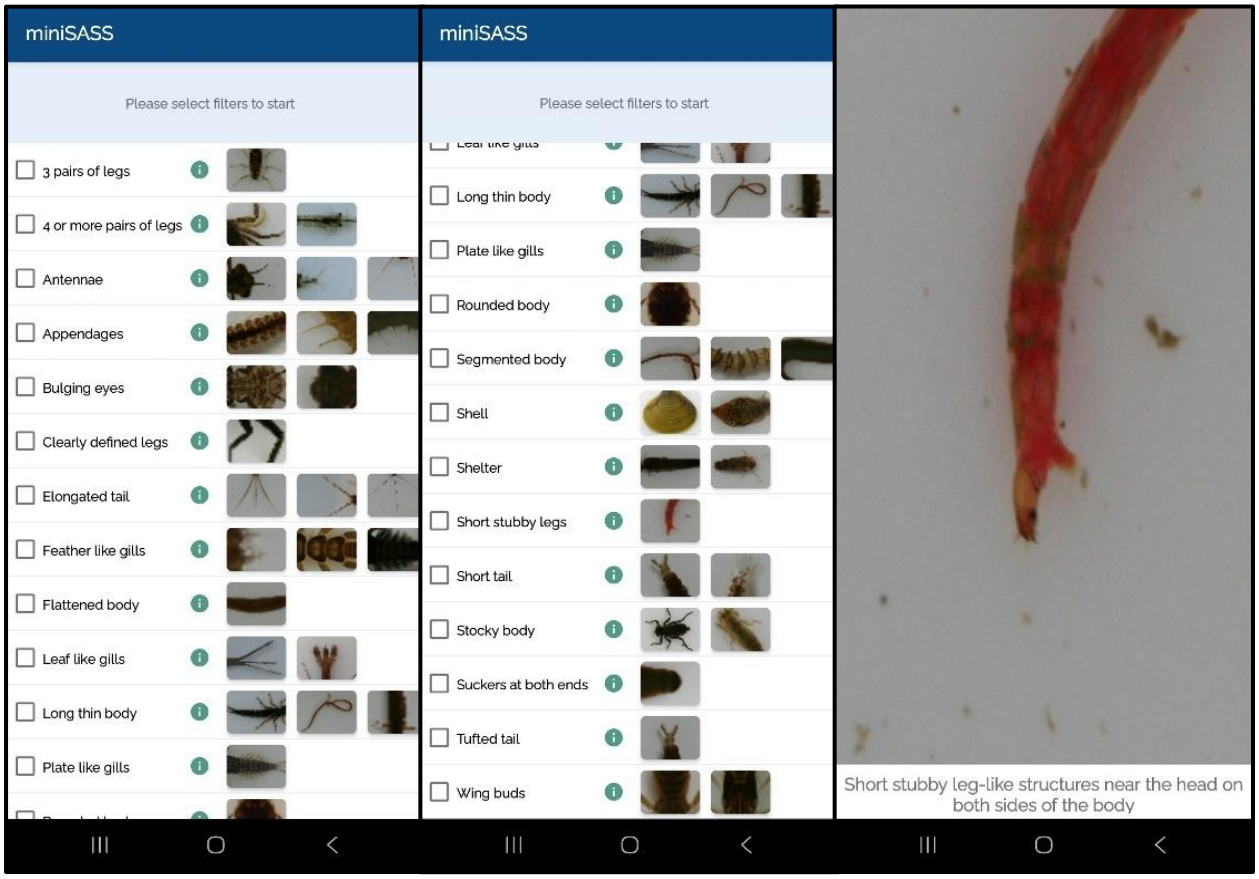


Figure 14: Illustration of the digital classification (left and middle) with an example of an identfier feature (right).

Once users have identified a feature, or multiple features, they can select the feature from the list, in this case – segmented body (Figure 15 left). Once the feature is selected all other features that do not accompany the selected feature is removed, for example, if segmented body is selected, three pairs of legs are removed from the list because none of the groups contain organisms that have both three pairs of legs and a segmented body. For clarification, a segmented body is defined as a body that is divided into equal parts, aside from the head or tail, this information is available in the FAQ section of the about page. Users can now select more features available from the list, in this case a long thin body (Figure 15 middle). Selecting multiple features provides users with possible groups listed at the top, they can select one of these groups are chosen to select additional features, in this case, an additional feature of appendages is selected (Figure 15 right). The combination of segmented body, long thin body and appendages is unique to true flies, and therefore the only group

available for selection is true flies. If the user is satisfied with the option provided, they can select it and add it to the list of macroinvertebrates found. This process is repeated until all macroinvertebrates are identified.

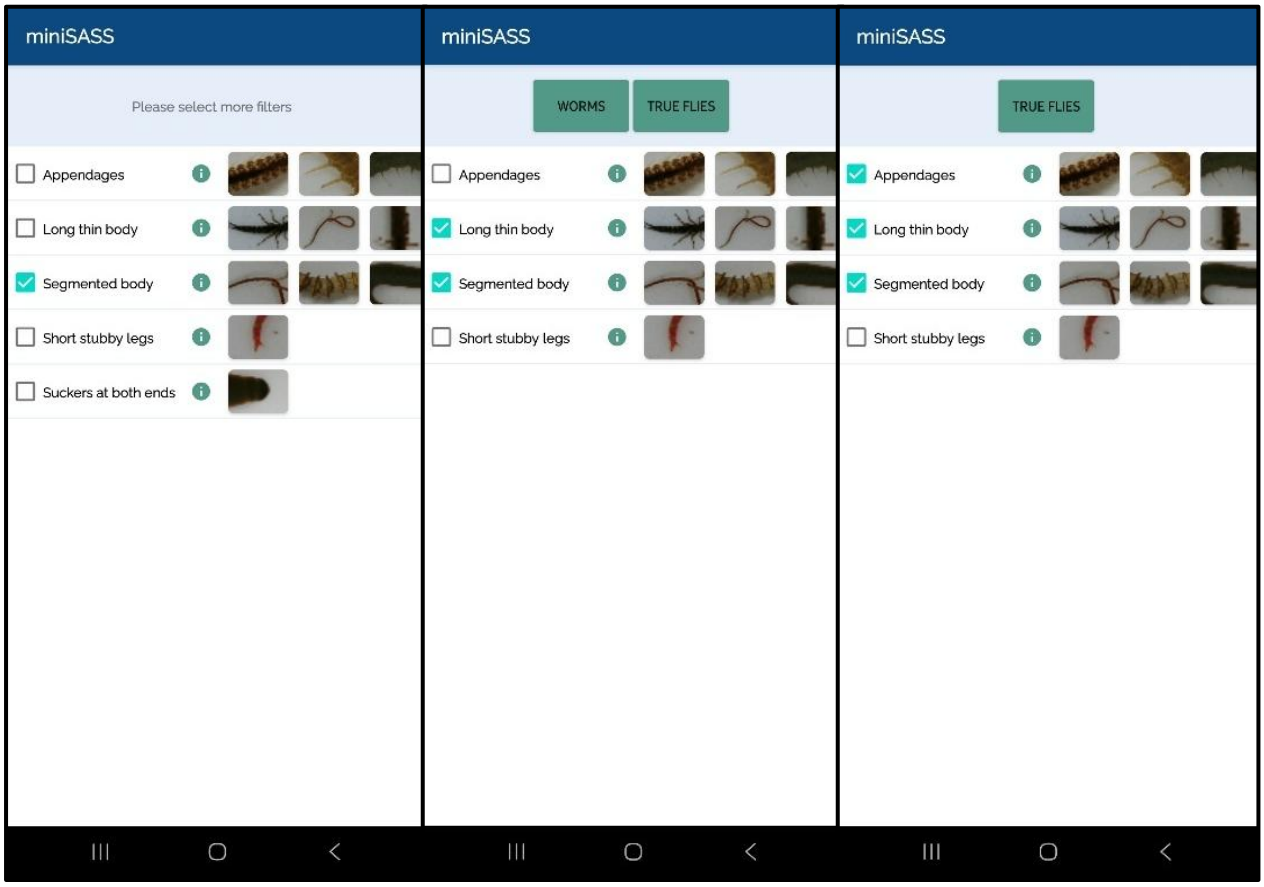


Figure 15: Example of the digital classification key in working.

2.3.4) miniSASS index scores

In Table 3 the user and ML generated miniSASS scores is given. In Table 4 the user and ML identification of groups is given, together with the accuracy of the ML algorthim. From Table 3, the scores for sites do not differ more than 0.3 units, except for the Mooi River Muiskraal site where the user score is 6.2 and the Ml score is 5.5, this is also where the ML algorthm has the lowest accuracy with 33%. The other site differing more than 0.3 units is the Wasgoedspruit 1 site, where the user score is 3.8 and the Ml score is 4.2, the ML accuracy is 67% for this site. The accuracy of the machine learning model to classify macroinvertebrates in the mobile application is above 60% for 9 of the 11 sites, with the Mooi River 4 sites havig 100% accuracy. Sites with accuracy lower than 60% are the Mooi River Carltonville road site where the aacuracy is 57%, although the miniSASS score does not differ much, this indicated

that although the scores are similar, the identification might not be accuracte and so addtional care should be taken when looking at generated scores with consuting the identfies groups. The other site with a accuracy below 60% is the Mooi River Groenwilgers site with an accuracy of 30%. The scores generated also differ with more than one unit, therbey indcating the misclassification of aquatic macroinvertebrates in this site leads to the large difference in generated scores. Given the low number of images used to train the algorthm and the scores geenrated from images taken in the field, the performance of the Ml algorthim to accurately classify aquatic macorinvertebrates was satisfactory.

Table 3: User scores versus machine learning (ML) scores as generated for miniSASS surveys using the mobile Application.

Site name	User score	ML score
Mooi River 3	4.75	5
Mooi River 4	5.14	5.14
Mooi River 5	4.33	4.4
Mooi River 8	4.33	4.4
Klerkskraal Dam wall	5.57	5.2
Mooi River Rysmierbult	4.5	4.6
Mooi River Muiskraal	6.2	5.5
Mooi River Carltonville road	4.83	4.75
Mooi River Groenwilgers	5.86	4.67
Wasgoedpruit 1	3.8	4.2
Wasgoedspruit 2	4.17	4

Table 4: Groups identified by users and ML prediction during surveying.

Site name	User groups	ML prediction	ML accuracy
Mooi River 3	Bugs & Beetles Flatworms Caddisflies Minnow Mayflies True flies Damselflies Damselflies Bugs & Beetles Snails/Clams/Mussels Crabs & Shrimps	Snails/Clams/Mussels True flies Caddisflies Minnow Mayflies True flies Damselflies Damselflies Bugs & Beetles Snails/Clams/Mussels Crabs & Shrimps	80%
Mooi River 4	Damselflies Snails/Clams/Mussels Worms Dragonflies Caddisflies Minnow Mayflies Crabs & Shrimps	Damselflies Snails/Clams/Mussels Worms Dragonflies Caddisflies Minnow Mayflies Crabs & Shrimps	100%
Mooi River 5	Snails/Clams/Mussels Minnow Mayflies Dragonflies Damselflies Bugs & Beetles True flies	Snails/Clams/Mussels Minnow Mayflies Dragonflies Minnow Mayflies Bugs & Beetles True flies	83%
Mooi River 8	Dragonflies Minnow Mayflies Worms Bugs & Beetles Snails/Clams/Mussels Damselflies	Dragonflies Minnow Mayflies Worms Bugs & Beetles Snails/Clams/Mussels Dragonflies	83%
Klerkskraal Dam	Caddisflies Worms Other Mayflies Snails/Clams/Mussels Damselflies Crabs & Shrimps Flatworms Crabs & Shrimps	Caddisflies Worms Minnow Mayflies Snails/Clams/Mussels Caddisflies Crabs & Shrimps Snails/Clams/Mussels Crabs & Shrimps	63%
Mooi River Rysmierbult	Snails/Clams/Mussels True flies Minnow Mayflies Damselflies Dragonflies Crabs & Shrimps	Snails/Clams/Mussels True flies Minnow Mayflies Minnow Mayflies Dragonflies Crabs & Shrimps	83%
Mooi River Muiskraal	Minnow Mayflies	Minnow Mayflies	60%

	Damselflies Other Mayflies Worms Caddisflies	Crabs & Shrimps Minnow Mayflies Worms Caddisflies	
Mooi River Carletonville road	Worms Caddisflies Minnow Mayflies Crabs & Shrimps Crabs & Shrimps True flies Bugs & Beetles	Worms Minnow Mayflies Minnow Mayflies Crabs & Shrimps Crabs & Shrimps Dragonflies Dragonflies	57%
Mooi River Groenwilgers	Snails/Clams/Mussels Minnow Mayflies Damselflies Flatworms Caddisflies Caddisflies Bugs & Beetles Bugs & Beetles Other Mayflies Damselflies	Snails/Clams/Mussels Crabs & Shrimps Crabs & Shrimps True flies Minnow Mayflies Bugs & Beetles Bugs & Beetles Bugs & Beetles Bugs & Beetles Dragonflies	30%
Wasgoedpruit 1	True flies True flies Damselflies Minnow Mayflies Worms Crabs & Shrimps	True flies True flies Dragonflies Minnow Mayflies Worms Crabs & Shrimps	83%
Wasgoedspruit 2	True flies Dragonflies Damselflies Worms Caddisflies Leeches	True flies Dragonflies Damselflies Worms True flies Crabs & Shrimps	67%

2.4) Discussion

2.4.1) miniSASS mobile application

The purpose of the miniSASS mobile application is to simplify the process of completing a miniSASS survey, by providing a single platform that houses the necessarily requirements for a survey, site creation, score calculation and organism classification. In addition to this the application features an about page where information of the importance of water quality monitoring is housed as well as ancillary and explanatory information on miniSASS and the macroinvertebrates used in the

survey process. Also, the about page lists safety precautions that users must follow to safely conduct surveys. The most important aspect of the about page is the training videos given. This serves to convey important information on how to conduct a miniSASS survey correctly, how to use the identification key and how to interpret the scores generated, and upload it to the website. It is not always possible to train citizen scientists on how to do miniSASS surveys and with this included in the application, the knowledge on how to conduct a miniSASS survey is more efficiently disseminated. The about page should be consulted before anyone proceeds with a miniSASS survey.

The landing page of the application has an important link to the training video on how to use the application and helps direct users to first consult the training media before conducting a survey. The map page is a simple interface where users can view their own sites as well as other created sites, however the map pages slow the application down since many sites must be displayed, the sites are already condensed in bubbles where the number of sites is displayed and as the users zooms in, the sites are displayed. Users can select a site and view a site score, however, users cannot navigate to the site on the website or access other information such as the images taken or ML classifications. Users also cannot add a survey to an existing site listed in the map, they can only add a survey to an existing site in the sites page.

The sites page is the main page of the miniSASS application. This page is compiled in a way to allow the user to input sufficient metadata, including site descriptions, site names, coordinates, date of sampling, river type, river name and notes, in addition to measured water quality. However, this information is not displayed in the map page when users view sites already created. After sites have been created, the site descriptions cannot be altered in the future and so the original description of the site remains with the original photo of the site and the date of creation. The site photo, description and date of sampling should be moved under the assessments section. Currently the assessment section does not display this information but does display the water quality measurements of each survey, the scores generated, the assessment photos and any additional notes. Once an assessment is created, users are also not able to change the assessment if they made a mistake, and a new assessment must be undertaken. During the capturing of macroinvertebrate images, the cropping of images works perfectly, the camera is automatically changed to macro mode to take better images of small objects. After the data is captured and organisms

identified the scores are generated, which takes a while to complete since the machine learning model is also classifying at this point.

Aside from the changes needed in the application it still offers an easy to use and quick way of conducting a miniSASS survey with what may be thought of as relatively minimal training. The mobile application is sufficient to conduct a miniSASS survey and succeeds in disseminating knowledge regarding water quality monitoring, aquatic macroinvertebrates and miniSASS scores to the user without the need to attend training or need any prior knowledge of miniSASS and water quality monitoring. The application provides a complete tool to educate users in the importance of water quality monitoring whilst providing them with a tool to conduct a survey and take part in the generation of sites scores for early pollution detection. Thereby users contribute to the throughput of data generated and stored on the miniSASS website.

2.4.2) Machine Learning (ML) model

The machine learning model was trained using 13 000 images of thirteen groups of macroinvertebrates as used in miniSASS, roughly on thousand images of each group was used for training, validation and testing. The model was trained using a Convolutional Neural Network (CNN). CNNs work well for image recognition through convolution of layers and pooling of features in the convoluted layers. This process is repeated until enough features are extracted in the images to accurately classify macroinvertebrates to group level. In total 512 layers were used as identifiers to classify macroinvertebrates. The training accuracy was 94.35% and a validation accuracy of 95.54%. This implies the CNN was successful in extracting features to classify macroinvertebrates during its training phase, it extracted the correct features to use as identifiers. The final classification accuracy during testing was 74% on unseen data. This is below a desired classification accuracy, however given the low number of images used to train the model this is satisfactory and shows CNNs can successfully classify aquatic macroinvertebrates.

With regards to the confusion matrix, the model has difficulty distinguishing between damselflies and dragonflies, this is possible due to the high morphological similarity in these organisms. Another problem is bugs and beetles, where only 50% of the time the model was accurate in classification, the reason for this low accuracy is the high morphological diversity within this group. Organisms within this group can either be in the adult or larval stage and hence have high morphological diversity. Beetles are also

one of the most diverse groups of organisms on the planet and identifying features to accurately classify and include all possible variations of aquatic beetles is difficult. In addition to bugs and beetles, true flies are also a problem group, this group also has high morphological diversity and looks like beetle larvae.

To combat this classification issue, it is recommended that the model be retrained on more images, preferably more than 10 000 per group but also including enough images of each group to accurately represent the high morphological diversity within groups. Additionally, it is recommended that classification starts at a higher taxonomic level, for example family level, where the images of organisms are grouped into families, this will lower the morphological diversity and improve the accuracy of the machine learning model. Thereafter the classification of organisms is at family level where all families belonging to the same group are then combined to represent that group. This will alleviate the classification inaccuracy and also begin to lay a groundwork for using machine learning to classify macroinvertebrates for SASS5, since families are used. Therefore, using family level identifications to train the model and then grouping those families into the 13 groups in miniSASS the accuracy of classification of scores used in miniSASS will be improved as the reliability of the machine learning score also improves and thereby removing the bottleneck for manual verification of site scores.

2.4.3) Digital classification key

The digital classification key was designed to increase the identification accuracy of aquatic macroinvertebrates. miniSASS score generated by users must be reliable when comparing to the ML model, the purpose of the ML model was to provide a tool for automatic verification of site scores and not to replace the user generated score. The digital key also serves as an educational tool where users can better understand the morphological variation of aquatic macroinvertebrates. The key was designed using 22 identifiers. Each macroinvertebrate group has a unique combination of these identifiers. For example, crabs and shrimps, have clearly defined legs, four or more pairs of legs, and antennae. Crabs and shrimps are not considered to have a shell, since they have an exoskeleton. Using these three identifiers and selecting them in the digital key, this group is displayed for selection. The key is successful in displaying each group using its unique identifiers and therefore using this digital key therefore simplifies the classification of macroinvertebrates, however, it does require the use of

the application. The images provided next to the identifiers provides the user with a visible representation of what to look for to ease the identification process, additionally, the digital classification key descriptions are given below the image as well as in the FAQ of the about page. It is recommended these identifiers be carefully revised to ensure a unique combination of features for each group, and to possibly add or remove features from the list to improve classification accuracy.

2.5) Summary

The miniSASS mobile application serves as a digital tool that allows users to more easily conduct a miniSASS survey, to disseminate knowledge to the public on macroinvertebrates as biological indicators of water quality and to introduce the use of machine learning to automatically classify macroinvertebrates. The application includes a home page, a map page, a field site creation page and an about page that in combination provides a fully functional and robust tool for biomonitoring using aquatic macroinvertebrates. The machine learning model included in the application was trained on 13 000 images of macroinvertebrates and is successful in classifying macroinvertebrates accurately to group level and can therefore serve as an automatically verifier of site scores. The miniSASS mobile application is therefore a tool that enhances the data capturing, analysis, data throughput and data storage of biological surveys using aquatic macroinvertebrates, whilst improving knowledge dissemination to the public and providing training tools and media, as well as introducing machine learning as a method of automatic verification of macroinvertebrates.

CHAPTER 3: DIATOM INDEX DEVELOPMENT

3.1) Introduction

3.1.1) Origin of biological indices

Throughout human history, the recognition of aquatic pollution has been primarily based on sensory indicators such as odour and visible deterioration in water quality. Early societies understood the health risks associated with consuming contaminated or stagnant water, often avoiding such sources due to the evident association with illness and mortality. Flowing or running water was generally regarded as safer for consumption, although its quality could not be guaranteed (Angelakis *et al*., 2023; Valenti, 2024). The development of organized sanitation systems is commonly attributed to the Romans, who engineered an extensive network of aqueducts and sewers designed to convey human and animal waste away from urban centres, thereby reducing exposure to pathogens and improving public hygiene. (Deming, 2019; Angelakis *et al*., 2023).

Pollution, in general became a noticeable issue during the industrial revolution in Europe, where large numbers of industrial factories were built and operated by burning coal (Hanlon, 2020). Air and water pollution became increasingly noticeable and large number of people flocked to cities and towns, often living in very poor conditions with little or no sanitation, which led to many people falling ill and dying from exposure to contaminated air and water, cholera outbreaks were not uncommon during the industrial revolution (Davenport *et al*., 2018). The quality of water declined steadily and was no longer fit for human consumption. People only acted when waterways were destroyed by pollution and waterways became unfishable and a general hazard to human health (Fisheries Preservation Association, 1868; Lenders, 2017). It is here, in the early 19th century where people started to understand aquatic pollution and sought to create a solution for its detection and mitigation. Events such as the “Great stink” in the Thames River caused politicians to act against such issues (Halliday, 1999). People noticed the water turning black and all kinds of fish and aquatic organisms dying. In the early 20th century, scientists sought biological indicators of aquatic pollution and started using bacteria, fish, aquatic macroinvertebrates and algae as indicators of aquatic pollution (Kolkwitz and Marsson, 1908; Thienemann, 1910).

The first index to detect aquatic pollution was developed by Kolkwitz and Marsson (1908) and was called the saprobic system. They grouped rivers into different zones based on the amount of organic pollution observed or known to be present. They divided rivers into three zones, oligosaprobic, mesosaprobic and polysabrobic, based on which organisms were present, and which were absent. And this became the first biological index of water quality (Kolkwitz & Marsson, 1908). In 1955, Pantle and Buck developed the first quantitative index for water quality assessment. Kolkwitz and Marrson's index was qualitatively based on presence and absence data whereas Pantle and Buck 1955 proposed a quantitative formula to calculate a saprobic value based on the tolerance values of diatoms, aquatic macroinvertebrates, higher plants, protozoa and bacteria present. Therefore, it was a start to diatom indices but not a true diatom index since it incorporated multiple groups of organisms in its calculation. Pantle and Buck 1955 assigned tolerance values to organisms between 1 and 4 based on where they were more likely to be found. If an organism was mostly found in clean water, a value of 1 was assigned and likewise if an organism was found mostly in organically polluted water, a value of 4 was assigned. Intermediate organisms were assigned values between 1 and 4. Based on the abundance of organisms found, a simple formula using weighted averaging was used to calculate a saprobic score for a particular site (Pantle & Buck, 1955). This method was groundbreaking progress in using aquatic organisms and indicators of organic pollution, however, it lacked taxonomic specificity and needed further refinement.

The approach was further refined by Zelinka and Marvan in 1961, when they proposed a new method of calculating an index score by introducing an abundance and indicator weight. The indicator weight reflects how reliable an organism is as an indicator, thereby refining the formula and offering a more robust calculation by down weighting less informative taxa and upweighting more informative taxa which reflect ecological state more closely. They also not only used diatoms but included higher plants, aquatic macroinvertebrates and other algae (Zelinka & Marvan, 1961). Zelinka and Marvan (1961) were also more taxonomically specific than Pantle and Buck, thereby further refining the index. The formula proposed by Zelinka and Marvan in 1961 served as the backbone for the development of most diatom indices used today, however, this index was still used only for organic pollution. Sládeček, in 1973 and onwards, contributed to the robustness of the saprobic systems by, throughout decades of research,

standardising indicator values for hundreds of taxa, including many organisms of different groups. Also, diatoms became the dominant indicator group in this index, leading to the first true diatom indices developed by Descy and Lange-Bertalot in the late 1970s by applying the formula of Zelinka and Marvan (1961) exclusively to diatoms (Descy, 1979). Lange-Bertalot himself did not create diatom indices but contributed significantly to species level identification of diatoms whereby sensitivity and tolerance values could be assigned to newly described taxa to improve the robustness of diatom indices (Lange-Bertalot, 1979; Bate *et al*., 2004).

3.1.2) History and working of diatom indices

From Kolkwitz and Marsson (1908) to Zelinka and Marvan (1961) the saprobic system was the principle biotic index used for biomonitoring, the system was refined over time and included many different groups of aquatic organisms as indicators of organic pollution. Building on the work of Zelinka and Marvan (1961), Descy (1979) developed the first true diatom index. He used diatoms because of their high diversity and abundance in river systems as well as their rapid response to nutrients and organic enrichment. They also preserve well in samples and can be permanently archived for future studies. These few reasons are, in part, why diatom indices have been so successful in their application. Descy assigned tolerance and sensitivity values to taxa commonly found in European rivers to create a diatom-only index for the detection of eutrophication and organic pollution (Descy 1979).

In the early 1980s, Coste developed the IPS index (Indice de Polluosensibilité Spécifique) based on Lange-Bertalot's taxonomy and expanded the original list created by Descy to include over 2000 taxa. The IPS included a larger number of taxa with more refined and standardized species tolerance and sensitivity values (Coste in CEMAGREF, 1982) and became a cornerstone in water quality monitoring in Europe for the detection of organic pollution and eutrophication. The IPS methodology contributed to the development of the Commission of the European Communities (CEC) diatom index, which was first proposed in the late 1980s and published in the early 1990s for application across European rivers (Taylor & Cocquyt, 2015). During this timeframe, the IPS was developed and improved to include more taxa and a standardized list of sensitivity and tolerance values, expanding on that of Descy in 1979. During the final development and dissemination of the CEC diatom index, the IPS methodology was incorporated into the framework and adopted as part of

standardized water quality monitoring in European rivers. Later, indices based on IPS/CEC approaches were integrated into the European Water Framework Directive to support ecological status assessment (Descy & Coste, 1991). The IPS still serves as the most widely used diatom index in Europe and most of the world. In countries where regionally calculated indices were developed, they were often compared to the IPS as the diatom index standard.

Since the development of the IPS, many diatom indices were developed following the approach of Zelinka and Marvan in 1961. Some of the most well-known indices are the Generic Diatom Index (GDI), developed in the early 1990s, as well as the Trophic Diatom Index (TDI) and the Biological Diatom Index (BDI), developed in the mid-1990s. The GDI follows the same calculation as the preceding indices but only focuses on genus level identification. This makes the index easier to use, however, because of the loss of taxonomic resolution, the ecological response is less accurately determined compared to species-level identification. The TDI, developed by Kelly and Whitton (1995) was created to detect nutrient enrichment, especially phosphorus released from wastewater treatment works, using a curated list of easy identifiable taxa. The BDI, developed by Lenoir and Coste (1996) serves as the national routine index of France and is used in large scale monitoring networks in the country. During the late 1990s and early 2000s, Rott developed trophic diatom indices to detect eutrophication instead of organic pollution, which became a critical inclusion in European directives (Rott *et al*., 1997; Rott *et al*., 1999; Rott *et al*., 2003).

The work of Taylor (2007) tested the application of these European indices in South African rivers to determine if they could be used as a biomonitoring tool in South Africa. Taylor (2007) was successful in his application of the European diatom indices in highly impacted waters, the IPS was especially successful. Harding and Taylor (2011) later included endemic South African taxa in the IPS index framework and included tolerance and sensitivity values for over 30 native and endemic taxa, thereby creating the first diatom index used in South Africa, the South African Diatom Index (SADI) (Harding & Taylor, 2011). The SADI is the most recent development in diatom indices in South Africa for riverine water quality monitoring, and no other riverine index has been proposed since. Indices were developed for the detection of AMD in wetlands in South Africa (Riato *et al.,* 2018).

The development of diatom indices requires the calculation and or calibration of species optima (sensitivity values) and tolerance ranges that are used in the calculation of the final index score. Most indices use the same methods and or a combination of methods. In the early development of diatom indices, such as those developed by Descy (1979) and Coste (1982), the calculation of species optima was based on the observed presence along an environmental gradient and the frequency at which they occurred, which is an empirical calculation of species optima. Additionally, species optima were also determined by expert knowledge, though years of observing diatoms in different environments and different ecological states. While the IPS remains widely used and now includes over 2 000 characterized taxa, literature suggests varying levels of calibration across taxa and that some optima likely remain derived from traditional observational methods rather than full WA re-calibration (Coste in CEMAGREF, 1982). The BDI, also developed by Coste (Lenoir & Coste, 1996) builds on the IPS framework, however, the BDI included more species and uses multiple stressors, including nutrient enrichment, salinity, habitat change and alkalinity to monitor water quality across French monitoring networks (Taylor & Cocquyt, 2015). Similarly, the GDI offers genus-level identification as a simplified alternate to the IPS developed from earlier datasets (Coste & Ayphassorho, 1991). The TDI (Kelly and Whitton, 1995) did not use empirically derived optima but rather WA to calculate species optima along a phosphorus gradient, specifically to detect eutrophication. The TDI also includes a %PTV value which is the percentage of valves counted that represent taxa tolerant to nutrient enrichment and serves as an excellent proxy for eutrophication in addition to the TDI score.

Other ways of calculating species optima for environmental parameters include Weighted Averaging using Partial Least Squares (WA-PLS) (ter Braak, 1993), Generalized Logit Regression models (GLR) and ordination techniques (Factor analysis, Correspondence Analysis (CA) and Canonical Correspondence Analysis (CCA)) (Smol & Stoermer, 2010). However, of all these methods, WA remains an efficient way of calculating species optima as it sits in a type of goldilocks zone of balance of accurately calculating optima whilst only needing a few occurrences of those species in sites sampled (ter Braak, 1987). WA is still the best representation of the unimodal distribution of diatom taxa over an environmental gradient, especially rare species which occur at the extremes of environmental gradients or do not reach

as high abundance in diatom communities compared to other taxa. WA is simple to explain, it is essentially an average calculation but with the requirement of a weight, which is normally the relative abundance of a species. In the context of diatom indices, WA aims to calculate a specific value of an environmental parameter but uses the abundance of a certain taxon as a weight. Taxa with high abundance at low values of the parameter pull the average downward, whereas taxa abundant at high values pull it upward, allowing inference of the taxa's optimum. Tolerance values can then be calculated as the standard deviation of the species' weighted occurrences around that mean optima (Shin *et al*., 2022). However, if the optima are at the extreme ends of the environmental gradient, the standard deviation often extends into negative values for the environmental gradient, which is ecologically nonsensical.

Another way to calculate tolerance values is using quantile regression, where species tolerance values are estimated by modeling quantiles across an environmental gradient (Cade & Noon, 2003). However, quantile regression requires larger datasets and computational models. Abundance-weighted quantiles provide a simplified yet robust alternative, in which the tolerance range of a species is determined across the environmental range it occupies, without using the environmental parameter as a predictor. Instead, species abundances are used as weights, giving greater influence to sites where the species is most abundant. This approach calculates species tolerance ranges in a practical and straightforward way while remaining robust, providing a realized ecological range that aligns with biomonitoring practices (Cristóbal *et al*., 2014: Solomon *et al*., 2025).The sites where species occur with higher abundance are weighted more and the lower 10% and 90% quantiles are calculated for the species distribution. The difference in the values for environmental parameters between the 90% quantile and the 10% quantile serves as the tolerance range for the species, essentially capturing 80% of the occurrences. Therefore, tolerance values never extend into negative environmental parameter values and span the environmental niche where 80% of the species occurs. By combining WA for optima with quantile-derived tolerance ranges, it is possible to better characterise species-response curves, especially for rare taxa, or those that occur at the extremes of the environmental range.

Another way of calculating species optima is using Generalised Logit Regression (GLR). This method involves three main concepts, gaussian curves, logistical

transformation of data and non-linear regression to calculate values used in the fitting of the gaussian curve (Smol & Stoermer, 2010). Using gaussian curves in ecology to fit species responses to environmental gradients is a common technique as many species have a unimodal, bimodal or sigmoidal response across an environmental gradient, making gaussian curves ecologically realistic (ter Braak, 1987). Logistical transformation of data refers to the alteration of species abundance data. In Smol and Stoermer (2010), they explain that GLR is mostly applied to presence absence data, however it is not appropriate to transform presence/absence data on a logarithmic scale. Therefore, using relative abundances of species is better for log transformation, however, species absences are still present, and a workaround is necessary for sites where species do not occur. It is acceptable to use a small value (e.g. 0.0001) for a zero abundance to be able to log transform species data. However, this produces large negative values in the dataset used to calculate the optima, tolerance and intensity (high of abundance) to fit the gaussian response curve. Therefore, in the present study it is proposed that only presence data for log transformation be used only if the taxon occurs in more than 10-15 sites with a relative abundance greater than 10%, partly based on Stevenson et al. (2008). This will ensure a better statistical calculation of optima and tolerance. Non-linear regression refers to the calculation of the species optima, tolerances and height of the curve (max abundance) assuming the species in question does not respond linearly to changes in an environmental variable but rather unimodally (Jamil *et al*., 2014).

GLR is useful for calculating species optima and tolerances, however only if species occur in many sites under a high relative abundance (Smol & Stoermer, 2010). Otherwise, using modeling techniques on species that are uncommon can be unnecessarily complex and runs the risk of overfitting (Breiner *et al*., 2015). Therefore, WA works better in most cases, but GLR and WA can potentially be combined. GLR can be used to calculate species optima for highly abundant and frequently occurring taxa whilst WA can be used to calculate optima for taxa that are rare and not as common or high in abundance in the diatom community (ter Braak & Looman, 1986; Jamil *et al*., 2014). However, the complexity of GLR might hinder its application in calculating species optima when WA works just as well and is a simpler method (ter Braak & Looman, 1986). Therefore, WA sits in compromise zone for species optima calculation.

Another method to derive species optima is from expert knowledge, given the IPS used this method and is highly accurate, it stands to reason that it is a good choice for species optima calculation. Additionally, expert knowledge on diatom taxonomy helps alleviate the misidentification of many species when assigning new values of species optima. Therefore, combining WA, GLR and expert knowledge is a good fit to calculate optima for many species, using the appropriate method for each species.

Another method used in WA-PLS, which is weighted averaging using partial least squares. This method is not as useful for calculating species optima but is more widely used to reconstruct past environmental conditions using diatom community structures (ter Braak, 1992). As WA-PLS is not commonly used for calculating species optima, it extracts the best fitting components using WA to more accurately determine what past environmental conditions were.

It is noteworthy to mention using ordination to determine species optima (Canonical Correspondence Analysis (CCA)) (ter Braak & Verdonschot, 1995). This method is not directly used to calculate species optima, it does give an excellent indication to species environmental relationships and incorporates the multifaceted and complex relationships that govern diatom community composition. This method can be used to confirm calculated species optima, as WA only uses a single metric to calculate optima and does not incorporate the complexity of environmental dynamics in the shaping of diatom community structures.

3.1.3) Interspecies and environmental relationships of diatoms

The dynamics and structure of diatom communities is difficult to understand. As with any aquatic community, it is difficult to observe diatom communities and the change in the community structure *in situ*, therefore, infering changes in the community based on changing environmental pressures becomes the only way to gain insight in the diatom community structure (Patil & Anil, 2005; Dalu *et al*., 2022). Snapshot observations allow for a determination of the current structure of a diatom community as it is represented in three-dimensional space, however observing changes in the community structure is within this space, in real time, not possible.

Diatoms have a complex relationship with the environment. They are largely ubiquitous, meaning they occur in every continent of earth and in every habitat type. However, whether the same species occur ubiquitously is debatable. Diatoms are

essentially photosynthetic ecosystem engineers, they colonise and shape a microenvironment through their different species traits (unicellular, colonial, motile, attached, etc.) (Hudon & Legendre, 1987; Rimet & Bouchez, 2012). They serve as the basis of the food web and sustain a complex community of biological species in many aquatic ecosystems (Béres *et al*., 2023; Stenger-Kovács & Béres, 2024).

In terms of ecological occupation, diatoms can be divided into categories based on the niche, or physical substrate they occupy. Diatoms that are unattached and free-floating are considered planktonic, whilst those that physically attached are called benthic diatoms (Passy, 2007). Within the division of benthic diatoms, there are those that occupy and live in sand (episammic), rocks and hard surfaces (epilithic), mud (epipelic) and those that attach to or are associated with aquatic plants (epiphytic). Planktonic diatoms are usually centric diatoms (rounded in shape) and do not possess a raphe, they have physical adaptations such as spines and spikes and large surface to volume ratios to prevent them from sinking and falling away from light (Round *et al*., 1990; Fragoso *et al*., 2018).

Benthic diatoms are largely represented by attached and motile diatoms that possess a raphe or a rimoportula by which they attach to the surface. A raphe is a longitudinal slit along the diatom valve that is a highly complex system of fissures where a mucilage substance (extra polymeric substances) is secreted that allow the diatom to glide along the surface (Round *et al*., 1990). A rimoportula is a structure that allows a diatom cell to attach to a surface at one of its apices and a such is located on the ends of the diatom valve as a single protrusion (Round *et al*., 1990). Additionally, diatoms can be either colonial or unicellular. Unicellular diatoms are attached to a surface either through their rimoportula, apical pore fields, or by means of the raphe. Colonial diatoms are attached to each other by means of interlocking spines or spikes and attached to the substrate by means of a mucilage stalk (Higgins *et al*., 2002; Nakov *et al.*, 2015). Colony formation in diatoms is often promoted when two-dimensional surface space becomes fully occupied and nutrient competition increases, in such cases, an extension into three-dimensional structures (e.g. stalks, mucilage matrices, extended colonies) may provide competitive advantages for nutrient uptake and light exposure (Mitchell *et al*., 2013; Wang & Tang, 2022).

In addition to the life forms and adaptations that diatoms must have to occupy specific niches within their environment, they can also be facilitating to one another.

Essentially, when diatoms colonise a new environment, the species that occupy the niche alter the environment either physically or chemically in a way that is beneficial for other diatom species or functional groups to occupy or to lose occupation of a niche (Vanelslander *et al*., 2009). As such, in the absence of pollution or detrimental impacts, diatom communities shape themselves and move toward a climax community structure where all niches are occupied and facilitation is actively shaping the diatom community succession.

With the presence of environmental pressures or pollution, the diatom community will alter in such a way that only those species adapted, either morphologically or physiologically to the new environmental condition will be able to compete for nutrients, space and light (Glevitzky *et al*., 2025). Therefore, the diatom community, at its climax, represents the environmental conditions that shape its structure (Stenger-Kovács *et al*., 2023). Diatoms respond rapidly to environmental changes and stressors and have short generation times, which allows the community to quickly adapt to the new conditions (Masouras *et al*., 2021).

The diatom community structure at a given location is therefore determined by the physical habitat available, the niche occupancy, the facilitation of species, and the water quality (mainly determined by geology and in part the trophy associated with the changing river continuum concept). It is therefore important to establish the structure of reference communities that form under specific conditions. With the addition of pollution or toxic substances, the diatom community will change and will be represented by those species able to withstand and even thrive in such conditions. Diatoms have undergone extensive adaptive radiation, and many diatom species specialise in a specific niche under specific environmental stressors (Godhe & Rynearson, 2017).

However, relying on morphological features to determine if a diatom taxon is tolerant to osmotic stress, nutrient enrichment or organic pollution is not a viable option as they all have different physiological adaptations to environmental change (Li *et al*., 2023; Stenger-Kovács *et al*., 2023; Bayramova, 2024). Physiological adaptations, such as gene expression under certain environmental conditions can help diatoms cope better to the environment in addition to morphological adaptations. Diatoms species within the same genus can have different physiological adaptations to handle environmental stress. Gene expression to synthesise or regulate nutrients transporters in the cell can

determine the ability of a species to withstand nutrient enrichment (Chen *et al*., 2018; Kang & Runearson *et al*., 2019; Dell'Aquila *et al*., 2020). Taxa with a high affinity for nitrate transporters for example can survive better in nutrient poor conditions because the uptake of nutrients, however, limiting, is better than taxa with a low affinity for nitrate. Similarly, the affinity for ammonium, phosphate and silica also plays a role in the survivability of a diatom in changing environmental conditions (Durkin *et al*., 2016; Kang & Runearson *et al*., 2019; Matsui *et al*., 2024). Also, ion transporters and osmolyte regulating genes play a role in the adaptations of diatoms toward osmotic stress and imbalance, and antioxidants enzymes, Extracellular Polymeric Substances (EPS) genes and ABC transporters can help cells cope better in organically polluted conditions (Su *et al*., 2024; Steele *et al*., 2014).

Therefore, using morphological, mechanical and physiological adaptations to determine if species are tolerant to specific environmental change is difficult and extremely complex, with molecular techniques becoming more accessible, in the future gene sequencing can be an invaluable tool for the calculation or determination of species optima and tolerance ranges. However, using weighted averaging to determine species optima across an environmental change is highly effective and remains the best option for index calculation at the present.

The distribution of diatom taxa follows, for the most part, a unimodal distribution along an environmental gradient with a single peak (Jamil *et al*., 2014). The community structure of diatoms is mostly determined by environmental factors such as substrate type, niche availability, biotope availability, nutrient availability, inter- and intraspecies relationships, as well as anthropogenic impacts such as pollution (Kelly & Whitton, 1995; Passy, 2007; Rimet & Bouchez, 2012; Göthe *et al*., 2013; Shibabaw *et al.*, 2021; Keck & Kahlert, 2019). Pollution, herein defined as detrimental changes to the environment by modification of water quality parameters which induces stress in the survival of diatom taxa, is the primary influence for the change in the diatom community structure outside natural factors, studies show that nutrient enrichment, and effluent discharge strongly correlate with shifts in diatom assemblage (Shikwambana *et al*., 2021; Masouras *et al*., 2021; Dalu *et al*., 2022). Under optimal water quality conditions, diatom assemblages form a climax community where diatom distribution is influenced by geology, niche availability, biotope availability and inter- and intraspecies relations, however, when environmental stress is introduced, diatom abundance will be affected

regardless of these aforementioned factors, and the community structure will be altered accordingly to favour the taxa able to withstand and compete under such externally induced changes (Chonova *et al*., 2019; Lu *et al*., 2020; Várbíró *et al*., 2020). Therefore, a principal factor influencing the abundance of diatom taxa within a naturally changing community assemblage is the deterioration or improvement of water quality within the system. The deterioration of water quality can be induced through salinization, pollution (AMD, organic, chemical, thermal, etc.) and nutrient enrichment (Riato *et al*., 2018; Donato-R *et al*., 2022; Stenger-Kovács *et al*., 2023).

The development of new diatom indices requires the use of large and reliable datasets that include as many as possible sites, taxa and accurate water quality measurements that reflect environmental change. For example, the Swedish phosphorus diatom index (PDI-SE) was developed using data from 820 stream sites (Kahlert *et al*., 2023). However, it is often the case that such datasets are scarce and regular gaps exists that can diminish the accuracy of diatom optima and tolerances calculated, and consequently diatom index results (Bate *et al*., 2004; Tison-Rosebery *et al*., 2023) . The improvement and testing of diatom indices require extensive fieldwork and the establishment of routine monitoring sites in different bioregions in South Africa, this also allows the identification of additional environmental and anthropogenic stressors to be included in the revision of diatom indices, as well as to gain insight into the distribution of species not previously recorded and not included in the calculation (Malan *et al*., 2005; Dalu & Froneman, 2016; Otto, 2018). In so doing the information housed in datasets used to calculate index scores is improved.

It is therefore important to consider the intra- and interspecies relationships in diatom communties as well as the environmental factors that govern their physiological responses when developing new indices. Additionally, creating and contributing to large datasets that house these environmental measurements and diatom community structures is important to accurately and reliably calculate new regional indices for routine monitroing and water quality resource management.

3.2) Methodology

For now, the use of large datasets is sufficient to test and compare various methods for the calculation of species optima and tolerances and consequently, diatom index scores. The size of a particular dataset, in terms of number of sites recorded is important, however, the abundances and diversity of species recorded in those sites

determine the useability of such datasets. For riverine indices a threshold of species inclusion was set at an abundance greater than 10% of the community composition and a presence in more than 10 sites, this is partly based on the study by Stevenson *et al*. (2008), who set the presence for species abundance at more than ten sites for optima calculation.

Five datasets, curated in the South African National Diatom Collection were used to establish species optima. Samples were collected from the Vaal River, the Crocodile Marico catchment, as part of three iterations of the Orange-Senqu River Commission study (ORASECOM – whole of the Orange River Basin from source to mouth) and from study conducted for the State of the Rivers report for the Durban Metro (KZN), samples were collected over 20 years from 2003 to 2023 with a total of 280 discrete sites. This data was cleaned, the nomenclature standardised and then used to calculate species optima and tolerance ranges from environmental conditions for riverine indices. After the indices were developed, they were correlated with the IPS, as well compared with measured water quality.

Epilithic diatoms were sampled and processed according to Taylor (2007b), samples were processed with acid digestion using the hot $HCl/KMnO_4$ method with additional organic material removal by H_2O_2 if necessary. Samples were cleaned and dried on coverslips before permanent mounting on slides with Pleurax (Von Stosch, 1974). At least 400 diatom frustules were counted for each site and taxa were identified to species level using a Nikon 80i microscope under 1000x magnification with Differential Interference Contrast (DIC) optics.

After diatoms were counted and identified, the corresponding environmental variables measured, *in situ,* were used to calculate optima (μ) and tolerance (v) values of diatom species across an environmental gradient. Environmental variables chosen for optima and tolerance calculation were electrical conductivity (EC), dissolved inorganic nitrogen (DIN), and orthophosphate (PO_4^{3-}).These variables were selected based on the effect that osmotic changes, organic- and nutrient enrichment have on water quality and diatom abundances. Salinisation reduces the useability of water for irrigation, domestic use and potability and increases the electrical conductivity of water, thereby placing stress on diatom communities (Stenger-Kovács *et al*., 2023). Treating salinized water is costly and salinisation is a common occurrence in streams affected by anthropogenic impacts (Schuler *et al*., 2019; Kaushal *et al*., 2021).

Additionally, nutrient enrichment from nitrogen (nitrates, nitrites and ammonium) and phosphate (total phosphates and orthophosphate) contributes most to the increase in the trophic state of rivers in South Africa and is the primary cause for nuisance algae blooms and the aesthetic deterioration of water quality (de Villiers & Thiart, 2007; DWS, 2022a; Lukhele & Msagati, 2024). Organic pollution is common in South Africa with failing infrastructure and deterioration of WWTF (Herbig, 2019; Jiyane *et al*., 2025). The primary diatom index currently used in South Africa is a modified form of the IPS, this index is used to determine the trophic state of rivers, ionic composition and organic content (Prygiel & Coste, 1993). Therefore, using similar metrics to calculate species optima is suitable for the comparison of newly calculated indices to the IPS.

The following methods explore the creation of riverine diatom indices using three different methods to calculate optima and tolerance values (Weigted Averaging (WA), Generalised Logit Regression (GLR) and inferred knowledge), then comparing scaling the optima and tolerance values (1-5 and 1-3 respectively) versus using absolute values in the index calculation (Figure 16). Thereafter, comparing calculated optima and tolerance values in a weighted average calculation versus a sum product calculation in the index score, and finally two methods of scaling the final index score (min/max scaling versus linear regression) for comparison and interpretation (Figure 16).

As mentioned, the riverine indices will explore different methods of calculating species optima and tolerance values - weighted averaging, for species optima, and weighted quantiles for species tolerance values under electric conductivity (EC), dissolved inorganic nitrogen (DIN) and orthophosphate (PO_4^{3-}). Weighted quantiles are explored due to the limitations of calculating tolerance values from the standard deviation from the optimum point and due to the complexity of quantile regression. When using diatom optima to determine the WA standard deviation of the unimodal distribution that the values for tolerance can extend into negative values for the environmental variable, which is ecologically impossible (ter Braak & van Tongeren, 1995). Abundance weighted quantiles (WQ) use the difference between abundance distribution thresholds as the tolerance range for diatom taxa, in this study the 10% and 90% abundance thresholds. WA is compared to GLR for both optima and tolerance under the same environmental variables. Both WA and GLR is compared to

the method of inferring optima and tolerance values from expert knowledge, where the occurrence of diatom species was observed over a 20-year period under nutrient load, ionic composition and organic load, much like the development of the IPS and GDI. The motivation for using these different methods stems from the fact that not all species are equally abundant or common in sampling sites. WA can be used to more accurately infer optima and tolerances for rarer and common species whilst GLR theoretically works well for common species but not for rare species (ter Braak & Looman, 1986; Jamil *et al*., 2014; Smol & Stoermer, 2010). Therefore, the present study explores which method of calculation works better, and if there is any difference in the results generated and compare these methods to expertly inferred values.

Once the optima and tolerance values were calculated using WA (Shin *et al*., 2022), GLR (Smol & Stoermer, 2010) and obtained from expert knowledge, they were rescaled to fit a value between 1-5 and 1-3 for optima and tolerance respectively and substituted into the formula proposed by Zelinka and Marvan (1961); this formula uses weighted averaging to calculate index scores and is given in section (3.2.2). The scales for optima and tolerance values as well as the formula is widely used for the calculation of diatom indices and was used in the IPS and other popular indices. Therefore, to more accurately compare calculated optima and tolerance values to the IPS they must be substituted into the same formula and be scaled in the same way into a final index score.

Additionally, index scores were also be generated using absolute values, not scaled between 1-5 and 1-3 into the same formula, to determine if the results varied significantly. The optima and tolerance inferred from expert knowledge was only be used as scale values in the index calculation as this was the original form in which they were recorded. In comparison to the weighted average formula, the optima and tolerance values (absolute and scaled) was used to calculate index scores using the sum product (formula 1 section 3.2.2). Thereafter, these indices were scaled either using the global minimum and maximum (0-100) of index scores generated across all sites as well as linear regression (1 – 20), as used in the IPS. The trophic scales proposed by Shikwambana *et al.* (2021) are used to infer ecological classes for indices regressed linearly (Table 7).

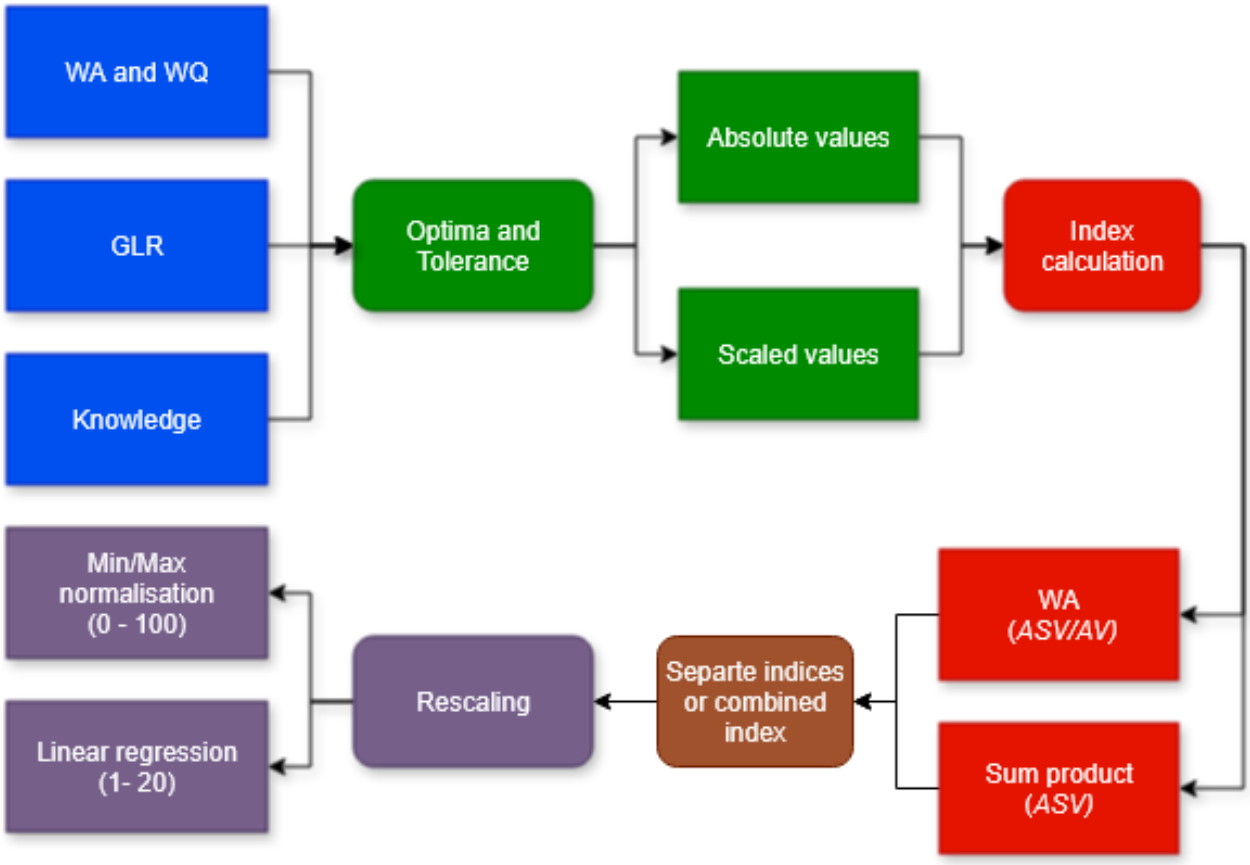


Figure 16: Workflow for diatom index calculation as followed by optima and tolerance calculation.

3.2.1) Calculation of species optima and tolerance values

3.2.1.1) Weighted Averaging (WA) and Generalised Logit Regression (GLR)

WA calculation only includes the influence of water quality on the abundance of a specific species and does not consider the influence of niche availability, biotype availability and inter- and intraspecies relationships directly (ter Braak & Verdonschot, 1995). However, the abundance of taxa at sites with varying degrees of pollution is used to determine their optimal environmental condition and therefore the influence of other factors is indirectly considered.

Optima and tolerances were calculated for the following environmental variables (Electrical Conductivity (EC), Dissolved Inorganic Nitrogen (DIN) and orthophosphate (PO_4^{3-})). EC is used as it represents the ionic compounds in water that conduct an electrical charge, and since osmotic stress is one of the most important factors determining diatom species distribution, it can be used with confidence (Patrick, 1977). The other environmental factors are nutrients (DIN and PO_4^{3-}), that are frequently associated with eutrophication and is a major concern for South African rivers (de Villiers & Thiart, 2007; Griffin, 2017; DWS, 2022a; Mararakanye *et al*., 2022; Lukhele

et al., 2024) and therefore, the use of these variables to determine species optima and tolerance ranges will help determine the degree of eutrophication in SA rivers. The IPS index scores applied in South Africa are based on species occurrences and primarily indicate levels of organic pollution (saprobity), while also accounting for ionic composition, trophic state and organic pollution (Coste in CEMAGREF, 1982; Taylor, 2007). Therefore, the use of EC, DIN and PO_4^{3-} in three separate, and combined calculations, will presumably be on par with the IPS index scores.

Before calculating species optima and tolerance values using WA, abundance weighted quantiles, and GLR, the datasets were combined and restructured. The calculation of all optima and tolerance values using WA were done manually in Excel without the use of software programs such as R, however, when calculating optima and tolerances using GLR, Python was used since the calculation is much more complex than simple weighted averages and quantiles. Species occurring in less than 10 sites with an abundance lower than 10% were omitted from the calculations. Optima and tolerance values for a total of 105 taxa were calculated for EC, DIN and PO_4^{3-}.

Calculating the species optima using WA was done by using the method of Shin *et al*. (2022). They used the following formula to calculate the optimum environmental condition for a species using weighted averaging (1) and using the standard deviation (2) as the tolerance:

$$\mu = \frac{\sum_{i=1}^{n} Y_{ik} X_i}{y+k} \quad y + k = \sum_{i=1}^{n} Y_{ik} \quad \textbf{(1)}$$

Where Y_{ik} is the abundance of the k-th species in the *I*-th region and X_i is the environmental variable in the *I*-th region (Shin *et al*., 2022). This formula uses the abundance of a single taxon across multiple sites as a weight to determine the average value for the environmental parameter where that species was found and uses this as the optimum environmental condition, or optima for that taxon.

The tolerance (σ) for species k within the *I*-th region can be calculated by using the following formula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (Xi - \mu k)^2 \, Yik}{\sum_{i=1}^{n} Yik}} \quad \textbf{(2)}$$

This formula uses the optima for a particular species, as calculated above, as a centre point and determines the standard deviation of the environmental variable from that point at which the species was recorded.

As mentioned, weighted quantiles were used to calculate species tolerance values instead of using the standard deviation. Weighted quantiles is based on the principle of cumulative abundance under increasing values of an environmental parameter, the cumulative abundance where a species reaches 10% abundance under all measured values for the environmental parameters serves as the lower threshold for the tolerance range, the cumulative abundance where a species reaches 90% of its total recorded abundance across increasing values for the environmental parameters serves as the upper threshold. Therefore, the weighted quantiles gives the 80% distribution range of a species under an environmental variable.

To use the WA formula for optima calculation, the data was structured to represent the relative abundance of a species across all 280 sites with the corresponding measured environmental parameters (EC, DIN and PO_4^{3-}). The environmental parameters were log transformed, using log 10, before calculating species optima and tolerances.

The values for optima and tolerances were taken as absolute values as well as scaled between 1-5 and 1-3 for optima and tolerance respectively. This was done to determine if converting species optima and tolerances to a scale value, as used in the original calculation of Zelinka and Marvan (1961) and the IPS, is appropriate and if the results differed from absolute values. The optima were calculated and converted to a scale using the following transformations:

Table 5: Species optima with corresponding ecological classes as converted from water quality ranges.

Ecological class	**Optimum**	**EC (µS/cm)**	**DIN (mg/L)**	**PO_4^{3-} (mg/L)**
Oligotrophic	5	<200	< 0.5	< 0.005
Meso-oligotrophic	4	200 - 350	0.5 - 1.5	0.005 - 0.015
Mesotrophic	3	350 - 600	1.5 - 2.5	0.015 - 0.05
Meso-eutrophic	2	600 - 1000	2.5 - 5.0	0.05 - 0.15
Eutrophic	1	> 1000	> 5.0	> 0.15

According to Table 5 species receive a score of 5 for EC when the calculated optima fall below 200 µS/cm, a value of 4 with a calculated optima between 200 and 350 µS/cm a value of 3 with an optima between 350 and 600 µS/cm, a value of 2 with a calculated optima between 600 and 1000 µS/cm and a value of 1 with an optima above 1000 µS/cm. A species receives a score of 5 for DIN when the calculated optima fall below 0.5 mg/L, a value of 4 with a calculated optima between 0.5 and 1.5 mg/L a value of 3 with an optima between 1.5 and 2.5 mg/L, a value of 2 with a calculated optima between 2.5 and 5 mg/L and a value of 1 with an optima above 5 mg/L. A species receives a score of 5 for PO_4^{3-} when the calculated optima falls below 0.005 mg/L, a value of 4 with a calculated optima between 0.005 and 0.015 mg/L, a value of 3 with an optima between 0.015 and 0.05 mg/L, a value of 2 with a calculated optima between 0.05 and 0.15 mg/L and a value of 1 with an optima above 0.15 mg/L. These conversions were based on Dallas & Day (2004) and DWAF (1996b) and have been altered slightly.

The calculated tolerance ranges for species were determined as the value between the 10% and 90% range of the species distribution. The 10% and 90% weighted quantiles of the environmental variable were calculated using species abundances as weights. For each species, the environmental values at all sites were sorted in ascending order, and the cumulative sum of relative abundances corresponding to the environmental variable value was computed. The environmental value corresponding to 10% and 90% of the cumulative weighted abundance was recorded as the lower and upper limits of the species' tolerance range respectively.

The tolerance values were scaled using the following transformations:

Table 6: Corresponding tolerance scale values as converted from the difference between the 10% and 90% distribution quantiles.

Parameter / Tolerance	**EC (µS/cm)**	**DIN (mg/L)**	**PO_4^{3-} (mg/L)**
3	value < 300	value < 1	value < 0.02
2	300 < value < 600	1 < value < 5	0.02 < value < 0.1
1	value >600	value > 5	value > 0.1

Species with larger tolerance ranges receive a lower score, therefore their influence on the community score is less. Common species that have a larger tolerance range for environmental parameters contribute less to the final index score, therefore rare species, which optima are calculated at the extreme ends of the environmental range, have a meaningful contribution to the index score when they are found.

According to Table 6, species with a tolerance range for EC less than 300 µS/cm receive a score of three, a calculated tolerance range between 300 and 600 µS/cm is converted to a value of two, and any tolerance values above 600µS/cm is converted to a value of one. Similarly, large tolerance ranges for DIN receive a lower score. Values calculated lower than 1 mg/L receive a value of three, values between 1 and 5 mg/L receive a value of two, and any tolerance value above 5 mg/L is converted to a value of one. Species with a tolerance range less than 0.02 mg/L PO_4^{3-} receive a score of three, those with tolerance ranges between 0.02 and 0.1 mg/L receive a score of two, and any tolerance range above 0.1 mg/l receives a score of one.

The transformations were loosely based on the environmental limits set out in Dallas and Day (2004) to define trophy levels based on the concentrations of nitrogen and phosphate in water whereby the trophic level increases. According to Table 5, when species optima calculated for an environmental parameter (EC, DIN and PO_4^{3-}) is low the corresponding inferred scaled value is high. Therefore, low species optima are converted to high scale values, this generates index scores that are high when the community optimum for the environmental parameter is low. Therefore, high index values correlate with low measurements for environmental parameters.

3.2.1.2) Inferred knowledge

The optima and tolerances values inferred from expert knowledge was determined through observation of diatom species under different environmental conditions over a period of 20 years. The samples used to infer these values were collected and processed as part of the South African National Diatom Collection (SANDC) housed at the NWU, South Africa. The curator for the collection, Prof Jonathan Taylor, has been studying diatoms in the field of ecology for over 20 years, and has contributed to the literature on the ecology of South African diatoms. He also modified the IPS index into the South African Diatom Index (SADI) by adding optima and tolerance values for over 30 endemic taxa. Therefore, his knowledge on diatom optima and tolerances are trusted and can potentially be reliably substituted to calculated index scores.

The complete dataset of species optima and tolerances as used in the IPS/SADI calculation was used to obtain species names. A total of 423 species is included in this dataset. The optima and tolerances for species for trophy, organic content, and ionic composition were determined, and their ecological description is also included (Appendix A.1 & A.2). These criteria define water pollution in South Africa and give a good proxy for environmental disturbance caused by nutrient enrichment, salination and organic pollution. Additionally, the IPS index scores correlate well with nutrient content, ionic load and organic content, and therefore determining species optima under these same criteria is important when comparing the inferred values to the IPS calculation. Furthermore, the IPS calculation only provides a single index score, where the inferred values are used to compute three separate indices, one for each variable, and combined into a final index score for comparison with the IPS.

The calculation of the final index score, as done in the IPS, will also be applied to these inferred optima and tolerance values. This maintains stability in the throughput of data and insodoing removes additional variables that may influence the comparison of final index scores. However, these inferred optima will also be substituted in the other proposed formulas, such as the sum product calculation.

3.2.2) Index calculation

The optima and tolerance values calculated using WA and inferred from knowledge, were substituted into two different formulas (formula 3 and 4), to calculate index scores for each environmental parameter as mentioned above; therefore, four separate indices were calculated for all sites using WA (scaled values in WA index calculation, absolute values in WA calculation, scaled values is a sum product calculation and absolute values in sum product calculation) and two different indices were calculated from the inferred optima and tolerances (scaled values in WA calculation and scaled values in sum product calculation). The indices were also combined into a multimetric index score by weighting the contribution of the individual indices (40% for EC, 25% for DIN and 35% for orthophosphate). The weighting of the individual indices are based on their relative ecological importance in South African rivers. Mining and agriculture are major contributors to water quality degredation, increasing the ionic load in rivers due to effluent inflow and irrigation return flows. Agriculture further contirbutes to nutrient enirhcment from nitrates and phosphate compounds used as fertiliser. Therefore, Electrical Conductivity (EC) is assigned the highest weight

because ionic strength and salinisation are major stressors influencing diatom community structures. Orthophosphate (PO_4^{3-}) was give a slightly lower weight as it is a key limiting nutrient that drives eutophication. Dissolved Inorganic Nitrogen (DIN) recieves a lower weight than orthophosphte because it generally has a smaller influence on diatom community structure compared to EC and PO_4^{3-}. This allignes with diatom biomonitoring studies in South Africa where nutrients and EC are primary drivers of diatom community structure (Taylor *et* al., 2007). Therefore, three metrics (EC, DIN and PO_4^{3-}) were combined into a single index score. In total, twenty different pathways of index calculation were explored.

Following the calculation of optima and tolerances for taxa, index scores were calculated using the following formulas:

$$\boldsymbol{I} = \sum_{j=1}^{n} a_j \, v_j s_j \quad (\mathbf{3})$$

$$\boldsymbol{I} = \sum_{j=1}^{n} a_j \, v_j s_j / \sum_{j=1}^{n} a_j v_j \quad (\mathbf{4})$$

The first is a sum product formula and the second is a WA formula where $\boldsymbol{a}_j$ represents the abundance of species ***j*** in a sample, $\boldsymbol{v}_j$ represents the tolerance value for species ***j*** in a sample and $\boldsymbol{s}_j$ represents the optimum environmental condition of species ***j*** in a sample. ***I*** is the calculated index value representing the weighted average sensitivity score of the entire diatom community.

The difference in these formulas is the use of weighted average to calculate the index score versus using a weighted sum. The first formula calculates the sum product of species abundances with their optima and tolerance values, whilst the second formula does the same calculation of the sum product but removes the weights of the abundance and tolerance values to obtain a weighted optima for the entire community.

Therefore, using formulas 3 and 4, the separate indices for each environmental variable is calculated, either using the absolute optima and tolerance values or those rescaled between 1-5 and 1-3 respectively. Thereafter, the separate indices were combined into a single index score using the following formulas:

$$MMI = (W_{EC} \times I_{EC}) + (W_{DIN} \times I_{DIN}) + (W_{po43-} \times I_{po43-}) \quad (\mathbf{5})$$

$$MMI = (W_{Ionic\ load} \times I_{Ionic\ load}) + (W_{nutrients} \times I_{nutrients}) + (W_{organic\ load} \times I_{organic\ load}) \quad (\mathbf{6})$$

Formula 5 is used to calculate index scores for optima and tolerance values calculated from WA, using the following weights for each variable: EC – 0.4; DIN – 0.25; PO_4^{3-} - 0.35. Formula 6 was used to calculate index scores for optima and tolerance values inferred from expert knowledge, using the following weights for each variable: Nutrients – 0.4; ionic load – 0.25; organic load– 0.35. These weights differ from the WA index weights, where EC had a higher weight and nutrients were lower, the focus of the WA index is more on overall trophic status and salinity. In the inferred knowledge index, nutrients (phosphorous and nitrogen compounds) get the highest weight, then organics and then EC. The idea is to make this index more sensitive to nutrient enrichment and organic pollution, while the WA index is better for overall trophic status and salinity.

3.2.3) Index scaling

After calculating the index scores using formulas 3 - 6, they must be presented in a way that is comparable. Scaling the index scores to be comparable is important when presenting the index score as a standalone value. Scaling was done using two methods, linear regression and minimum/maximum scaling. The outputs for formula 3 are large values since it is a sum product calculation, therefore regardless of using scaled values or absolute values, the final index value is large and scaling these values is best done using a minimum and maximum approach where the global maximum and global minimum of index scores is used to scale all values between 0 and 100, this reduces the large numbers into a interpretable scale value. Using formula 4 the outputs are weighted and therefore the range of the index value is the range of the optima values input into the calculation, whether it is absolute values or scaled values. Therefore, using minimum/maximum scaling is still appropriate, however, when using scaled values in formula 4, the output is a value between 1 and 5 as calculated in the IPS index. Therefore, this value is better interpreted on the same scale as the IPS (1-20) using the same formula for linear regression (Formula 8). Additionally, the trophic classes as defined by Shikwambana *et al*. (2021) are on a scale from 1 – 20, therefore linearly transforming the index scores to fit this scale seems appropriate for comparison with the IPS.

Scaling the index values using the global minimum and global maximum and linear regression between 1 and 20 is done using the following formulas:

$$Index_{min/max} = \frac{Index_S - Index_{min}}{Index_{max} - Index_{min}} * 100 \quad \textbf{(7)}$$

$$Index\ (scaled) = I(4.75) - 3.75 \quad \textbf{(8)}$$

Formula 5 represents the calculation using minimum and maximum values, where the index score is normalized between the global maximum and minimum, and multiplied by 100 to obtain a value between 0 and 100. Formula 8 represents the linear regression as used in the OMNIDIA v 5.3 software for the IPS calculation where the index value (I) is linearly regressed to a value between 1 – 20. This is different from multiplying the value with four to obtaining a value between 1 and 20 (Figure 17).

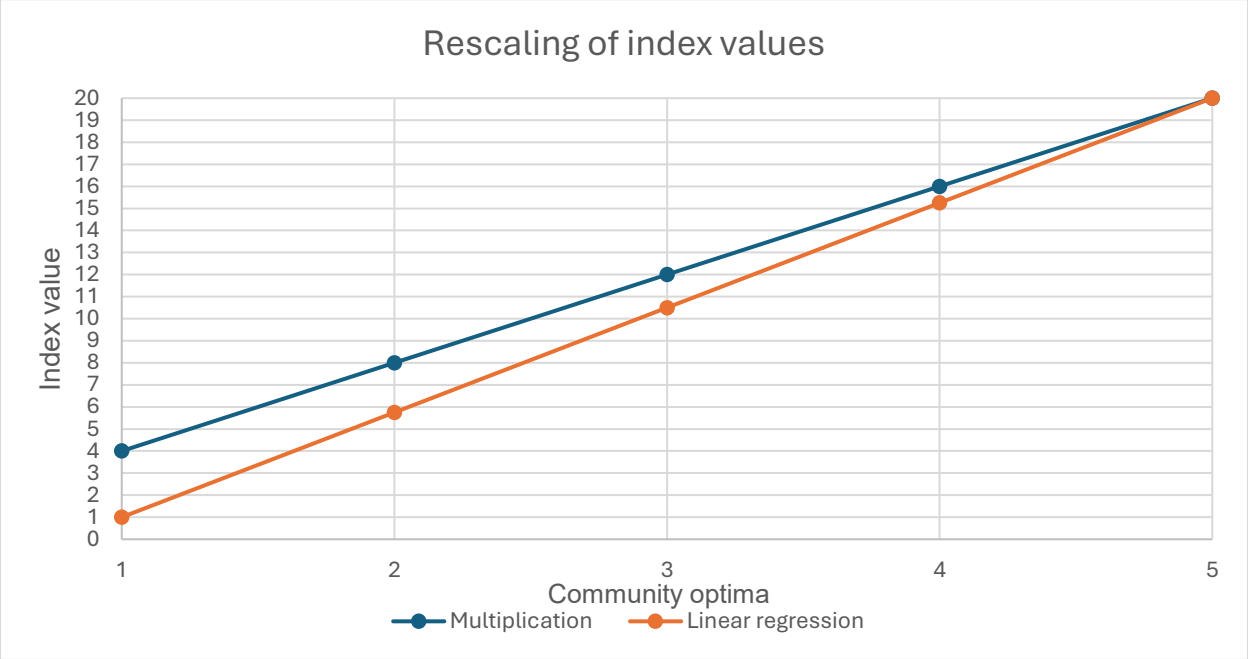


Figure 17: scaling of index values from a value between1 and 5 to a value between 1 and 20. The blue line represents multiplying the community optima by 4 to obtain a value between 1 and 20 where the orange line represents the linear regression formula as used in the IPS calculation.

In Figure 17, a clear distinction can be seen between the results of scaling optima using multiplication to fit values between 1 and 20, and those obtained through linear regression. With linear regression, the conversion of optima into a 1–20 range occurs more gradually. In contrast, when using multiplication, an optima value of 1 is converted to 4, whereas with the linear regression method, the optima remains at 1.

The combined and separately calculated indices were correlated with measured water quality using Pearson correlation and p-values to determine the relationship between index scores and water quality, before Pearson correlation was done, indices were checked for normality using the Kolmogorov-Smirnov test. Additionally, the indices

were correlated with the IPS index and the IPS index was also correlated with measured water quality. Redundancy Analysis (RDA) ordination plots were also created to illustrate the relationship of calculated indices to measured water quality.

The trophic levels in Shikwambana *et al.* (2021) were used to define threshold for index scores since the same thresholds are used when the IPS is applied.

Table 7: Ecological classes corresponding to IPS index scores and trophic levels.

Ecological class	Water quality	Trophic level	IPS score
A	High	Oligotrophic	>17
B	Good	Oligo-mesotrophic	15-17
C	Moderate	Mesotrophic	12-15
D	Poor	Meso-eutrophic	9-12
E	Bad	Eutrophic	<9

3.3) Results

3.3.1) Calculated species optima and tolerances using WA

Optima and tolerances ranges were calculated for 105 taxa using a combination of five datasets, combining 280 sites with species abundances, EC, DIN and PO_4^{3-}. The abundance of each species at a site was taken with the corresponding environmental variable and used to calculate a weighted average optimum according to Shin *et al*. (2022). The number of sites used differ for each species, where common species occurred in more sites than rarer species with lower dominance in the community. The tolerance range for each species was determined using the 10% and 90% abundance thresholds under each environmental parameter. Species optima and tolerances were converted to a scale between 1-5 and 1-3 respectively using Tables (5) and (6). Additionally, a composite or combined optima and tolerance range was determined using the same weights used to combine the separate indices into an average index (EC – 40%, DIN – 25% and PO_4^{3-} - 35%) (Table 8). This value was used to determine the combined trophic preference and tolerance class for each species. Each optimum was rescaled to a value between 1 and 20 and used to determine the trophic preference for each species Table (7). Similarly, tolerance classes were determined

for each species, these however, were divided differently than optima. Species with a tolerance below 1.5 were placed into the generalist class, species with tolerances between 1.5 and 2.25 were placed into the intermediate class and those with tolerances greater than 2.25 were placed in the specialist class.

Table 8: Calculated optima and tolerances for species under EC, DIN and PO_4^{3-}. Combined optima and tolerances with corresponding trophic preference and tolerance ranges (full names for species codes given in Appendix A.1 & A.2).

CODE	nr of sites	EC optima (uS/cm)	EC tolerance (uS/cm)	μ	v	DIN optima (mg/L)	DIN tolerence (mg/L)	μ	v	PO_4^{3-} optima (mg/L)	PO_4^{3-} tolerance (mg/L)	μ	v	Combined optima	Combined tolerance	Trophic preference	Tolerance class
ADEG	63	571.48	410.41	3	1	2.01	6.52	3	1	0.18	1.57	1	1	2.3	1	meso-eutrophic	Generalist
ADMI	140	183.64	413.391	5	1	0.14	0.43	5	3	0.02	0.04	3	2	4.3	1.85	oligo-mesotrophic	Intermediate
ADRI	15	158.33	87.75	5	3	0.5	0.78	5	3	0.06	0.21	2	1	3.95	2.3	oligo-mesotrophic	Specialist
ADSA	91	273.29	298	4	2	0.2	0.63	5	3	0.02	0.08	3	1	3.9	1.9	oligo-mesotrophic	Intermediate
AEXI	12	499.25	194.03	3	3	1.81	4.94	3	2	0.5	2.15	1	1	2.3	2.05	meso-eutrophic	Intermediate
AMMO	57	374.18	701.13	3	1	0.21	0.94	5	3	0.05	0.22	3	1	3.5	1.5	mesotrophic	Generalist
APED	116	517.29	719.04	3	1	0.38	3.05	5	2	0.05	0.93	3	1	3.5	1.25	mesotrophic	Generalist
AUGA	57	348.43	465.56	4	1	0.14	1.09	5	2	0.03	0.11	3	1	3.9	1.25	oligo-mesotrophic	Generalist
AUGR	85	643.21	556.5	2	1	0.49	2.23	5	2	0.1	0.24	2	1	2.75	1.25	meso-eutrophic	Generalist
AVEN	14	1152.97	828.93	1	1	4.91	8.16	2	1	0.42	0.99	1	1	1.25	1	eutrophic	Generalist
CAFF	17	331.93	536.29	4	1	0.13	0.54	5	3	0.01	0.02	4	2	4.25	1.85	oligo-mesotrophic	Intermediate
CINV	33	569.8	632.73	3	1	0.28	1.54	5	2	0.07	0.26	2	1	3.15	1.25	meso-eutrophic	Generalist
CKAP	33	244.01	355.38	4	2	0.2	0.2	5	3	0.03	0.04	3	2	3.9	2.25	oligo-mesotrophic	Intermediate
CMED	16	821.55	155.17	2	3	0.14	0.56	5	3	0.13	0.08	2	1	2.75	2.3	meso-eutrophic	Specialist
CMEN	69	884.84	895.46	2	1	0.59	8.7	4	1	0.16	0.86	1	1	2.15	1	meso-eutrophic	Generalist
CMLF	59	252.47	675.41	4	1	0.24	0.51	5	3	0.06	0.13	2	1	3.55	1.5	mesotrophic	Generalist
CPED	83	495.33	608.58	3	1	0.49	1.93	5	2	0.08	1.67	2	1	3.15	1.25	meso-eutrophic	Generalist
CPLA	151	391.04	634.46	3	1	0.35	2.09	5	2	0.06	0.16	2	1	3.15	1.25	meso-eutrophic	Generalist

CPLE	66	448.01	398.99	3	2	1.26	3.05	4	2	0.11	1.56	2	1	2.9	1.65	meso-eutrophic	Intermediate
CTGL	34	206.46	127.98	4	3	0.25	0.33	5	3	0.03	0.07	3	1	3.9	2.3	oligo-mesotrophic	Specialist
CTUM	27	216.65	358.71	4	2	0.16	0.18	5	3	0.03	0.04	3	2	3.9	2.25	oligo-mesotrophic	Intermediate
DCOF	13	545.65	269.93	3	2	1.64	1.87	3	2	0.1	0.39	2	1	2.65	1.65	meso-eutrophic	Intermediate
DCON	20	193.41	318.97	5	2	2.35	1.58	3	2	0.02	0.35	3	1	3.8	1.65	mesotrophic	Intermediate
DVUL	57	351.72	643.89	3	1	0.24	2.32	5	2	0.03	0.17	3	1	3.5	1.25	mesotrophic	Generalist
EMIN	13	397.92	178.39	3	3	0.44	2.28	5	2	0.02	0.01	3	3	3.5	2.75	mesotrophic	Specialist
ENCM	18	399.92	584.26	3	1	0.16	0.41	5	3	0.02	0.01	3	3	3.5	2.2	mesotrophic	Intermediate
ENLS	22	215.57	414.62	4	1	0.39	1.84	5	2	0.06	0.13	2	1	3.55	1.25	mesotrophic	Generalist
ENMI	59	215.93	470.58	4	1	0.21	1.7	5	2	0.03	0.21	3	1	3.9	1.25	oligo-mesotrophic	Generalist
EOLI	42	406.82	588.42	3	1	1.38	9.91	4	1	0.18	3.14	1	1	2.55	1	meso-eutrophic	Generalist
EOMI	25	524.4	577.5	3	1	0.92	2.94	4	2	0.04	0.13	3	1	3.25	1.25	mesotrophic	Generalist
EORC	27	643.17	420.01	2	1	0.49	1.39	5	2	0.13	0.09	2	1	2.75	1.25	meso-eutrophic	Generalist
ESBM	161	455.2	755.77	3	1	0.91	5.19	4	1	0.14	0.93	2	1	2.9	1	meso-eutrophic	Generalist
ESOR	12	453.93	82.03	3	3	0.13	0.08	5	3	0.05	0.02	3	2	3.5	2.65	mesotrophic	Specialist
FBCP	28	292.84	390.64	4	2	0.29	2.12	5	2	0.04	0.32	3	1	3.9	1.65	oligo-mesotrophic	Intermediate
FBID	26	157.28	210.32	5	2	0.37	0.58	5	3	0.04	0.05	3	2	4.3	2.25	oligo-mesotrophic	Intermediate
FBRE	18	398.34	401.05	3	1	1.94	2.85	3	2	0.25	0.91	1	1	2.3	1.25	meso-eutrophic	Generalist
FCAP	15	210.02	144.03	4	3	0.42	2.54	5	2	0.02	0.01	3	3	3.9	2.75	oligo-mesotrophic	Specialist
FCVA	40	262.69	352.22	4	2	0.2	0.36	5	3	0.05	0.1	3	1	3.9	1.9	oligo-mesotrophic	Intermediate
FSAP	83	368.08	587.65	3	1	1.29	3.91	4	2	0.09	0.31	2	1	2.9	1.25	meso-eutrophic	Generalist
FULN	17	235.46	691.22	4	1	0.26	3.37	5	2	0.03	0.13	3	1	3.9	1.25	oligo-mesotrophic	Generalist
GEXL	12	101.31	772.51	5	1	0.13	0.14	5	3	0.01	0.02	4	2	4.65	1.85	oligo-mesotrophic	Intermediate

GLST	12	118.19	179	5	3	0.21	0.24	5	3	0.03	0.03	3	2	4.3	2.65	oligo-mesotrophic	Specialist
GMIN	77	242.26	417.55	4	1	0.21	0.48	5	3	0.03	0.05	3	2	3.9	1.85	oligo-mesotrophic	Intermediate
GOMP	20	362	364.86	3	2	1.02	1.82	4	2	0.18	1.81	1	1	2.55	1.65	meso-eutrophic	Intermediate
GPAR	208	318.16	696.78	4	1	0.41	2.7	5	2	0.06	0.31	2	1	3.55	1.25	mesotrophic	Generalist
GPSA	17	1050.43	1104.22	1	1	5.79	7.35	1	1	0.61	0.88	1	1	1	1	eutrophic	Generalist
GPUM	89	327.62	500.13	4	1	0.63	3.56	4	2	0.08	1.33	2	1	3.3	1.25	mesotrophic	Generalist
GVNU	48	224.68	418.21	4	1	0.81	3.11	4	2	0.1	1.68	2	1	3.3	1.25	mesotrophic	Generalist
MAPE	103	436.65	592.89	3	1	1.04	4.08	4	2	0.19	1.18	1	1	2.55	1.25	meso-eutrophic	Generalist
MVAR	59	433.92	611.04	3	1	0.66	3.14	4	2	0.06	0.19	2	1	2.9	1.25	meso-eutrophic	Generalist
NACI	30	394	327.06	3	2	0.1	0.27	5	3	0.04	0.05	3	2	3.5	2.25	mesotrophic	Intermediate
NADF	26	590.96	206.73	3	2	0.29	1.41	5	2	0.08	0.11	2	1	3.15	1.65	meso-eutrophic	Intermediate
NAMP	73	502.65	452.84	3	1	4.01	7.95	2	1	0.53	3.32	1	1	2.05	1	eutrophic	Generalist
NANT	34	465.68	304.42	3	2	2.93	7.71	2	1	0.3	2.29	1	1	2.05	1.4	eutrophic	Generalist
NARV	26	398.92	517.96	3	1	2.46	8.02	3	1	0.3	2.77	1	1	2.3	1	meso-eutrophic	Generalist
NAVS	14	368.07	438.91	3	1	0.97	1.85	4	2	0.11	0.35	2	1	2.9	1.25	meso-eutrophic	Generalist
NCPL	65	832.74	1225.67	2	1	1.26	8.72	4	1	0.13	0.88	2	1	2.5	1	meso-eutrophic	Generalist
NCPR	53	311.4	268.32	4	2	0.19	0.41	5	3	0.03	0.12	3	1	3.9	1.9	oligo-mesotrophic	Intermediate
NCRY	29	349.51	778.35	4	1	0.36	3.56	5	2	0.04	0.09	3	1	3.9	1.25	oligo-mesotrophic	Generalist
NCTE	99	437.51	787.1	3	1	0.22	2.95	5	2	0.04	0.08	3	1	3.5	1.25	mesotrophic	Generalist
NDIS	52	324.2	607.28	4	1	0.15	0.5	5	3	0.03	0.05	3	2	3.9	1.85	oligo-mesotrophic	Intermediate
NERI	70	546.18	655.84	3	1	0.35	2.89	5	2	0.1	0.3	2	1	3.15	1.25	meso-eutrophic	Generalist
NFIL	34	848.33	1056.04	2	1	0.34	9.04	5	1	0.18	0.95	1	1	2.4	1	meso-eutrophic	Generalist
NFON	55	650.62	757.03	2	1	0.2	1.26	5	2	0.06	0.3	2	1	2.75	1.25	meso-eutrophic	Generalist

NGER	60	269.94	456.23	4	1	0.58	3.16	4	2	0.07	0.3	2	1	3.3	1.25	mesotrophic	Generalist
NGRE	85	609.12	600.88	2	1	1.8	2.46	3	2	0.16	0.27	1	1	1.9	1.25	eutrophic	Generalist
NIAR	20	604.24	407.57	2	1	0.3	3.66	5	2	0.05	0.04	3	2	3.1	1.6	meso-eutrophic	Intermediate
NIFR	196	485.03	546	3	1	0.49	2.93	5	2	0.09	0.49	2	1	3.15	1.25	meso-eutrophic	Generalist
NIGR	14	729.3	85.44	2	3	0.12	0.19	5	3	0.03	0.02	3	2	3.1	2.65	meso-eutrophic	Specialist
NINC	36	874.82	565.16	2	1	1.4	3.98	4	2	0.28	0.67	1	1	2.15	1.25	meso-eutrophic	Generalist
NIPU	29	406.68	266.45	3	2	0.1	0.27	5	3	0.03	0.04	3	2	3.5	2.25	mesotrophic	Intermediate
NLBT	74	399.52	632.03	3	1	0.25	1.73	5	2	0.06	0.14	2	1	3.15	1.25	meso-eutrophic	Generalist
NLIN	11	524.82	1446.43	3	1	0.36	5.76	5	1	0.03	0.05	3	2	3.5	1.35	mesotrophic	Generalist
NLSU	11	290.86	253.3	4	2	0.17	0.14	5	3	0.04	0.03	3	2	3.9	2.25	oligo-mesotrophic	Intermediate
NMEN	82	411.06	596.32	3	1	0.2	1.71	5	2	0.04	0.17	3	1	3.5	1.25	mesotrophic	Generalist
NMIN	18	361.2	839.78	3	1	0.15	0.38	5	3	0.07	0.13	2	1	3.15	1.5	meso-eutrophic	Generalist
NPAE	65	474.59	1236.26	3	1	0.43	8.92	5	1	0.08	0.87	2	1	3.15	1	meso-eutrophic	Generalist
NPAL	207	486.31	728.21	3	1	0.63	7.9	4	1	0.13	1.07	2	1	2.9	1	meso-eutrophic	Generalist
NPHO	18	334.51	262.19	4	2	0.79	1.43	4	2	0.1	0.98	2	1	3.3	1.65	mesotrophic	Intermediate
NRCS	53	662.11	543.31	2	1	0.18	1.26	5	2	0.08	0.23	2	1	2.75	1.25	meso-eutrophic	Generalist
NROS	98	432.89	569.31	3	1	0.51	4.4	4	2	0.08	0.59	2	1	2.9	1.25	meso-eutrophic	Generalist
NSHR	46	365.82	413.64	3	1	1.65	4.52	3	2	0.15	0.9	2	1	2.65	1.25	meso-eutrophic	Generalist
NSLT	29	353.65	454.51	3	1	0.89	1.71	4	2	0.06	0.21	2	1	2.9	1.25	meso-eutrophic	Generalist
NSYM	137	425.96	543.59	3	1	1.46	9.27	4	1	0.21	1.57	1	1	2.55	1	meso-eutrophic	Generalist
NTEN	45	596.13	787.17	3	1	0.6	3.69	4	2	0.04	0.04	3	2	3.25	1.6	mesotrophic	Intermediate
NVDA	12	215.07	280.96	4	2	0.38	0.35	5	3	0.05	0.01	3	3	3.9	2.6	oligo-mesotrophic	Specialist
NVEN	128	797.94	1156.34	2	1	0.79	8.68	4	1	0.16	0.84	1	1	2.15	1	meso-eutrophic	Generalist

NZSU	22	354.01	424.16	3	1	0.82	4.08	4	2	0.12	0.64	2	1	2.9	1.25	meso-eutrophic	Generalist
PDAU	16	524.95	874.86	3	1	1.28	1.96	4	2	0.06	0.08	2	1	2.9	1.25	meso-eutrophic	Generalist
PENG	36	458.4	390	3	2	1.09	2.6	4	2	0.02	0.1	3	1	3.25	1.65	mesotrophic	Intermediate
PLFR	44	608.41	442.89	2	1	1.56	4.72	3	2	0.06	0.08	2	1	2.25	1.25	meso-eutrophic	Generalist
POBG	25	257.21	185.14	4	3	1.43	1.59	4	2	0.06	0.15	2	1	3.3	2.05	mesotrophic	Intermediate
PTRO	18	315.41	601.64	4	1	0.85	2.68	4	2	0.05	0.21	3	1	3.65	1.25	mesotrophic	Generalist
RABB	50	599.61	479.8	3	1	0.98	2.49	4	2	0.16	1.93	1	1	2.55	1.25	meso-eutrophic	Generalist
RUNI	66	289.28	300.4	4	2	0.21	0.45	5	3	0.05	0.06	3	1	3.9	1.9	oligo-mesotrophic	Intermediate
SAGA	91	707	446.88	2	1	0.56	1.9	4	2	0.1	0.21	2	1	2.5	1.25	meso-eutrophic	Generalist
SBRE	30	683.37	414.47	2	1	0.82	2.13	4	2	0.09	0.19	2	1	2.5	1.25	meso-eutrophic	Generalist
SELI	12	848	610.49	2	1	2.35	0.51	3	3	0.12	0.09	2	1	2.25	1.5	meso-eutrophic	Generalist
SHAN	31	823.06	112.36	2	3	0.59	1.44	4	2	0.1	0.16	2	1	2.5	2.05	meso-eutrophic	Intermediate
SMST	57	437.41	369.55	3	2	1.42	3.94	4	2	0.2	1.98	1	1	2.55	1.65	meso-eutrophic	Intermediate
SPAV	16	599.27	455.79	3	1	0.21	1.61	5	2	0.09	0.24	2	1	3.15	1.25	meso-eutrophic	Generalist
SPUP	35	594.74	634.68	3	1	2.72	8.16	2	1	0.23	1.42	1	1	2.05	1	eutrophic	Generalist
SSEM	74	491.82	725.28	3	1	3.49	9.19	2	1	0.48	4.1	1	1	2.05	1	eutrophic	Generalist
TAPI	48	735.98	587.13	2	1	0.48	3.82	5	2	0.06	0.22	2	1	2.75	1.25	meso-eutrophic	Generalist
TPSN	18	598.38	516.84	3	1	0.15	0.49	5	3	0.07	0.21	2	1	3.15	1.5	meso-eutrophic	Generalist

Table (8) contains the Weighted Average (WA) optima and tolerances for 105 species as calculated for EC, DIN and PO_4^{3-}, the number of sites used to calculate these values are also included. The optima and tolerances are given as absolute values for each parameter and as scaled values. The conversion from the absolute values to the scale values was done according to Table (5) for optima and Table (6) for tolerances. These converted optima and tolerances values were combined into a single value according to the same weight conversions as used in the index calculation (EC – 0.4, DIN – 0.2 and PO_4^{3-} - 0.35).

The combined optima were placed into trophic preference according to the classes in Table (7) by using linear regression to convert the 1-20 value to a 1-5 value. The converted tolerances were placed into tolerance classes (3 - specialist, 2 – intermediate and 1 – generalist). Table (8) contains the optima and tolerances as used in the individual index calculation as well as the combined index calculation.

For EC, six species were classified as oligohaline based on calculated optima, corresponding to low electrolyte content (< 200µS/cm). Twenty-seven species were classified as oligo-mesohaline, corresponding to low to moderate electrolyte content (200 - 350 µS/cm (Figure 18). Fifty-one species were classified as mesohaline, corresponding to moderate electrolyte content (350 – 600 µS/cm). Nineteen species were classified as meso-euryhaline, corresponding to moderate to high electrolyte content (600 - 1000 µS/cm). Only two species were classified as euryhaline, corresponding to high electrolyte content (> 1000 µS/cm). Overall, most species are classified as mesohaline and oligo-mesohaline for EC (Figure 18 & 19).

For DIN, fifty-seven species were classified as oligotrophic (< 0.5mg/L) according to their calculated optima. Thirty-two species were classified as oligo-mesotrophic (0.5 – 1.5mg/L) (Figure 22 & 23). Ten species were classified as mesotrophic (1.5 – 2.5mg/L). Five species are classified as meso-eutrophic (2.5 – 5mg/L), and only one species was classified as eutrophic (> 5mg/L). Most species were classified as oligotrophic and oligo-mesotrophic for DIN (Figure 22 & 23).

For PO_4^{3-} , none of the species were classified as oligotrophic (<0.005mg/L). Two species were classified as oligo-mesotrophic (0.005 – 0.015mg/L). Thirty-eight species were

classified as mesotrophic (0.015 – 0.05mg/L) (Figure 21 & 22). Forty-four species were classified as meso-eutrophic (0.05 – 0.15mg/L) and twenty-one species were classified as eutrophic (>0.15mg/L). Overall most species were classified as mesotrophic and meso-eutrophic for PO_4^{3-}.

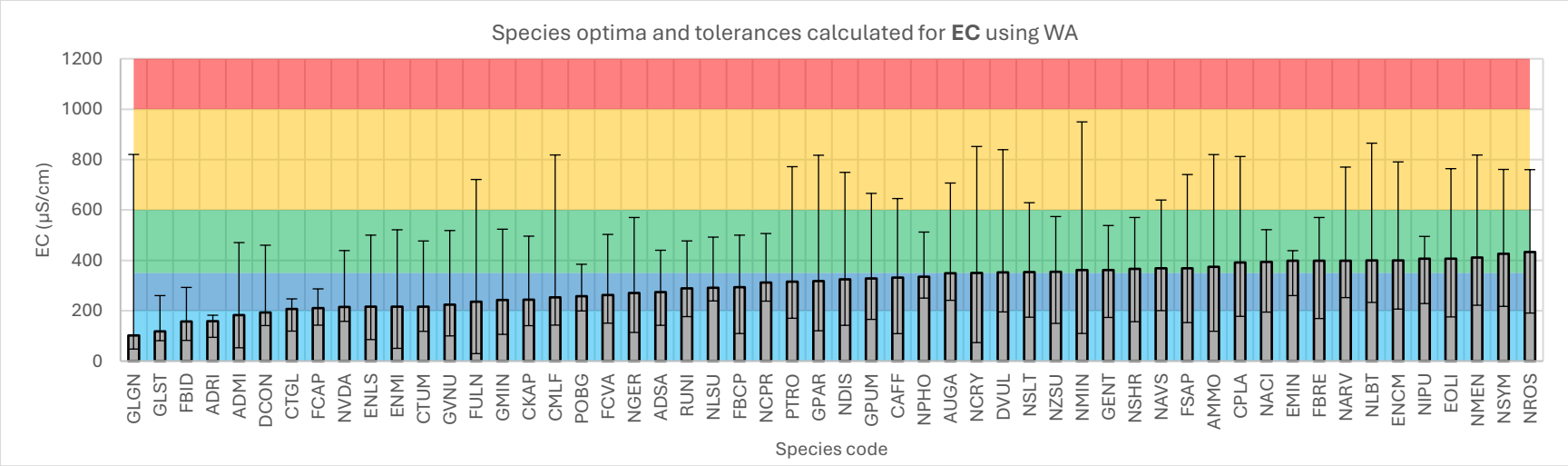


Figure 18: Optima (blue bars) and tolerance ranges (error bars) calculated for taxa for EC. Sections are colour coded based on Table (7) and represent ecological classes.

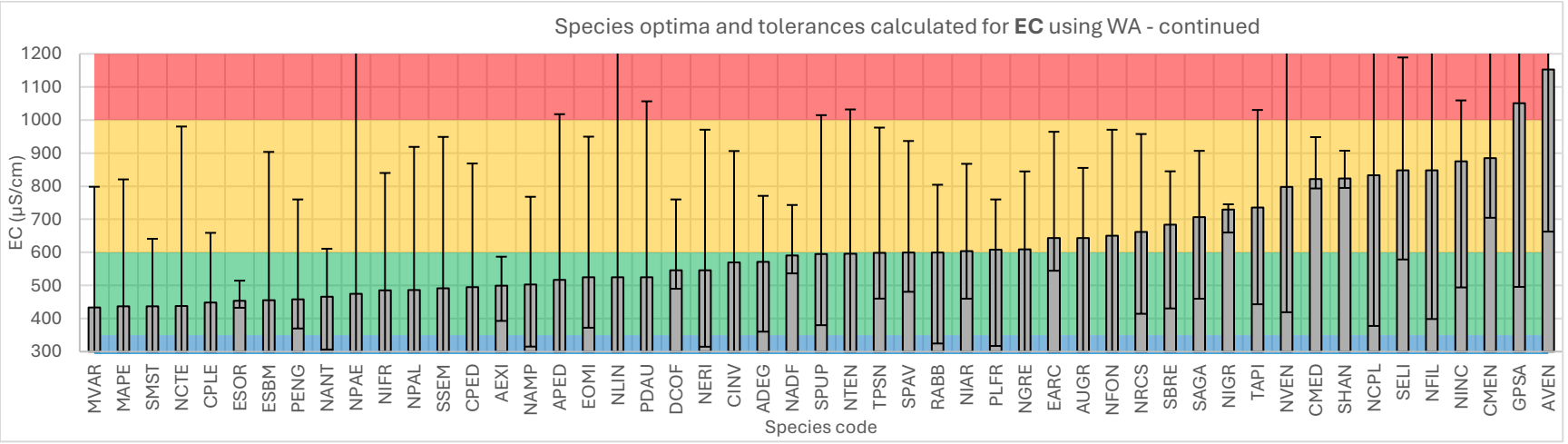


Figure 19: Optima (blue bars) and tolerance ranges (error bars) calculated for taxa for EC- continued. Sections are colour coded based on Table (7) and represent ecological classes.

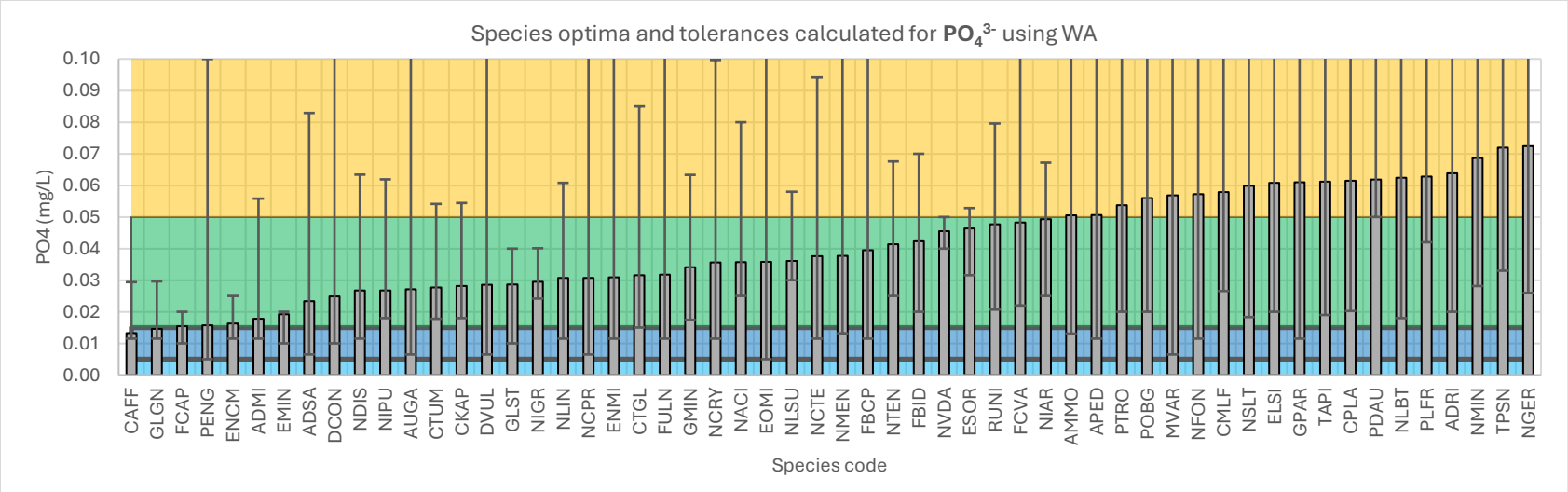


Figure 20: Optima (blue bars) and tolerance ranges (error bars) calculated for taxa for PO_4^{3-}. Sections are colour coded based on Table (7) and represent ecological classes.

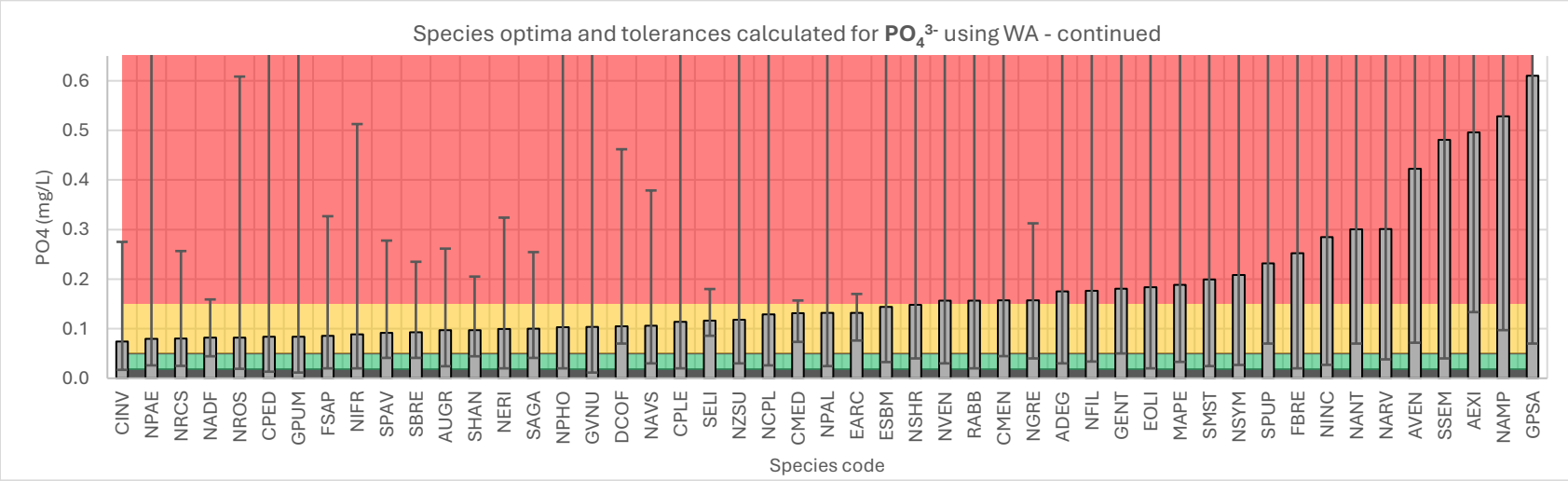


Figure 21: Optima (blue bars) and tolerance ranges (error bars) calculated for taxa for PO_4^{3-} - continued. Sections are colour coded based on Table (7) and represent ecological classes.

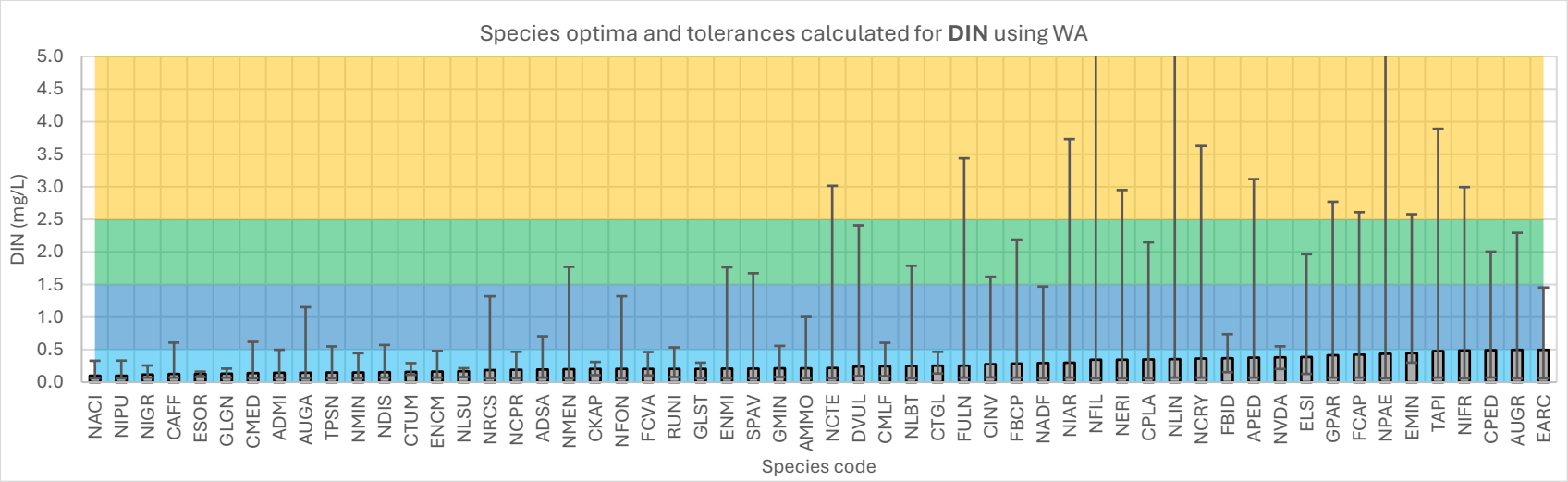


Figure 22: Optima (blue bars) and tolerance ranges (error bars) calculated for taxa for DIN. Sections are colour coded based on Table (7) and represent ecological classes.

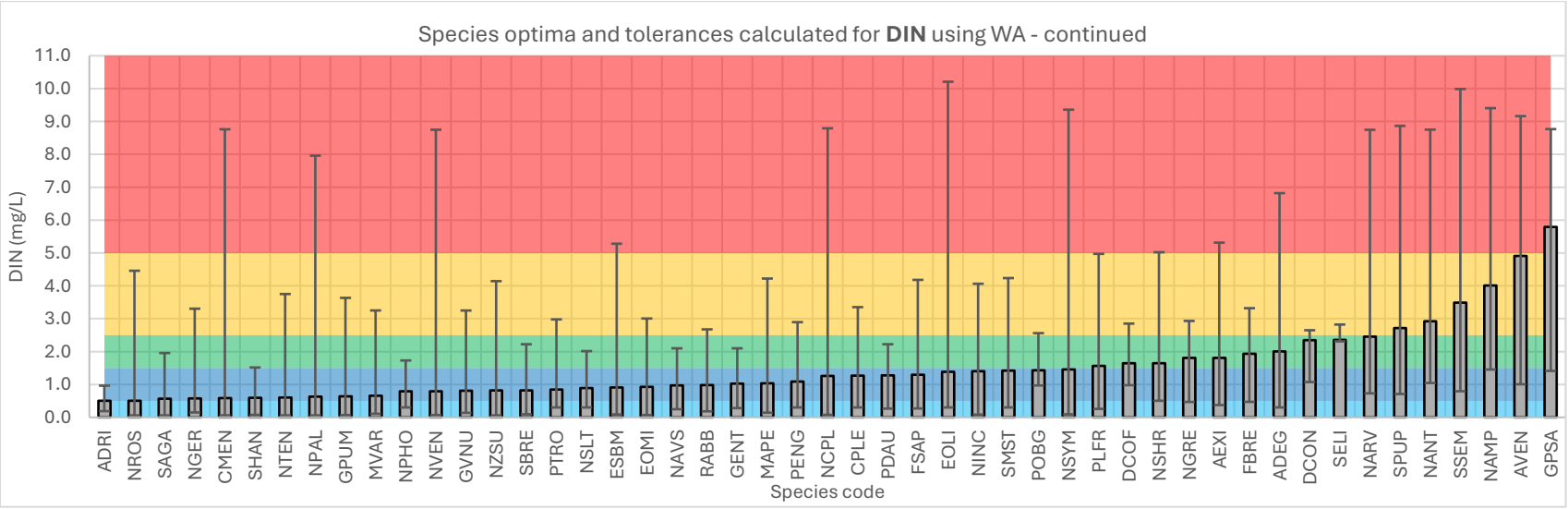


Figure 23: Optima (blue bars) and tolerance ranges (error bars) calculated for taxa for DIN - continued. Sections are colour coded based on Table (7) and represent ecological classes.

Figure 24 illustrates a normal distribution of species for each of the environmental parameters respectively (EC, DIN and PO_4^{3-}), reflecting the water quality of the rivers used to calculate the optima. Most species have growth optima at moderate electrolyte content, moderate to high PO_4^{3-} and low DIN.

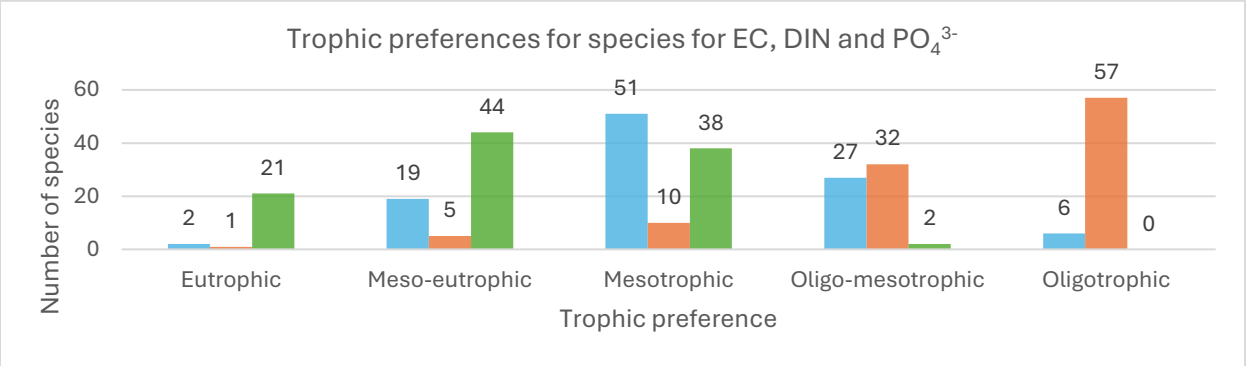


Figure 24: Trophic preference of species as determined from calculated optima for EC (blue bars), DIN (orange bars) and PO_4^{3-} (green bars).

Species are grouped into tolerance classes according to Table 6. For EC, species fall mostly into the generalist class (seventy-three) and secondly into the intermediate class (twenty-one). Only eleven species were classified as specialists under EC (Figure 25). For PO_4^{3-} species follow a similar classification to EC, with most species classified as generalists (eighty-four). Seventeen species were classified as intermediate, and only four species were classified as specialists (Figure 25). For DIN most species were classified as intermediates (Fifty-nine), eighteen were classified as generalists and twenty-eight were classified as specialists. Most species are generalist for EC and PO_4^{3-}, whilst for DIN, most species are intermediate.

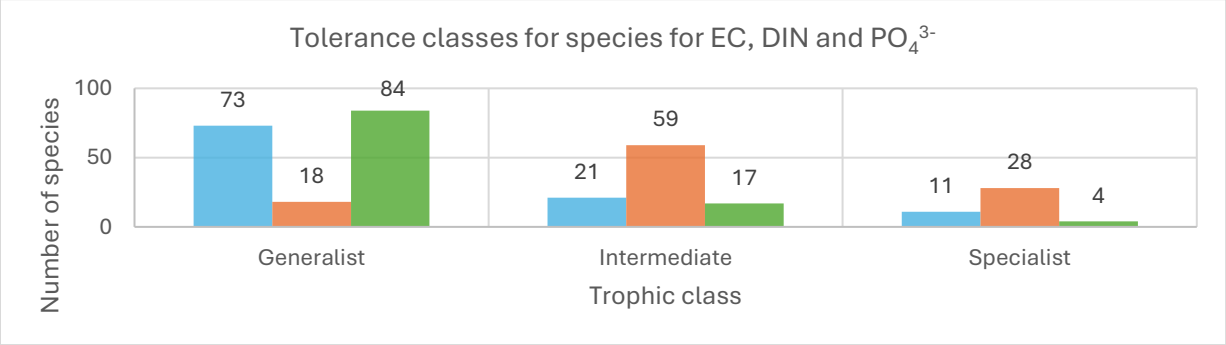


Figure 25: Tolerance classes of species as determined from calculated tolerance ranges for EC (blue bars), DIN (orange bars) and PO_4^{3-} (green bars).

3.3.2) Optima inferred using Generalised Logit Regression (GLR)

GLR was explored to estimate species optima and tolerances along the environmental gradient. Species exhibited sparse occurrences and clustered distributions along the gradient. Optima estimated from GLR were inconsistent across species. Weighted Averaging (WA) produced optima estimated for a substantially larger proportion of species and was therefore used for further analyses.

3.3.3) Optima and tolerances inferred from expert knowledge

The optima and tolerance ranges for species inferred from expert knowledge were divided into three separate categories: Trophic state (nutrient content), ionic load and organic content. These categories were chosen to reflect the common pollution effects present in South Africa such as agriculture (trophic level), salinization through irrigation (ionic load) and industrial pollution and wastewater treatment failure (organic content). Additionally, the IPS correlates well with nutrient content, ionic load and organic composition and therefore for better comparison of results these categories were chosen. Furthermore, the optima and tolerance scales are the same as the IPS and other categories; optima between 1 and 5 and tolerance ranges between 1 and 3. Optima and tolerance values for 423 taxa were determined together with the ecology of species. These values were inferred from observation over 20 years of research. These values are listed in Appendix A1.

3.3.4) Selection of final indices based on similarity and correlation.

Optima and tolerances used in the proposed index calculations were done using Weighted Averaging (WA) and inferred knowledge. Generalised Logit Regression (GLR) did not work due to ecological and statistical limitations. These optima and tolerances were used to calculate indices according to Figure (16), where different pathways of transforming the optima and tolerance were explored. After optima and tolerance were calculated they were either used as absolute values or scaled values in the index calculation. The indices were calculated according to section 3.2.2 using one of two methods, the WA calculation by Zelinka and Marvan (1961) (Formula 4) and a sum product calculation (Formula 3). After index calculation for each parameter, the indices were combined into a single value using Formulas 5 and 6. The individual and combined

index scores were scaled according to section 3.3.3 using one of two methods, minimum and maximum scaling on a scale of 0 – 100 (Formula 7) and linear regression on a scale of 1 – 20 (Formula 8). The selection of the final indices is based on the elimination of methods within this workflow.

Only eight of the calculated indices were chosen to present as final indices. The selection is partly based on the correlation between the individual indices for each parameter with each other and the correlation of the combined indices with each other, and partly based on the correlation of indices with the water quality parameters. Indices that follow fewer transformations have slightly better correlations with measured water quality variables, both for individual indices and combined indices (Table 9). These higher values may indicate slightly higher correlations with individual parameters and reflect the environmental gradient, however, might not reflect the broader ecological response as well as indices with more transformations. All indices were tested for normality using the Kolmogorov–Smirnov test, this test is appropriate for $n > 50$. All indices have values $p < 0.05$ and thus they do not deviate from normal distribution. Therefore, Pearson correlation was used for all indices with water quality parameters used to calculate the indices (EC, DIN and PO_4^{3-}). The possibility of intercorrelation is greatly decreased through the transformation of optima and tolerances from the water quality values and the further transformations during index calculation.

Table 9: Pearson's correlation coefficient of calculated indices with measured water quality (all values have significant correlations with $p < 0.05$). (KBI – Knowledge Based Index; WAI – Weighted average Index; S – scaled optima and tolerances; ABS – absolute optima and tolerances; SP – sum product calculation; MIN/MAX – minimum/maximum transformation; LIN – Linear regression).

Index	EC	DIN	PO_4^{3-}
IPS	-0.5527	-0.3761	-0.5306
S KBI-N LIN	-0.6208	-0.2608	-0.4759
S KBI-O LIN	-0.5698	-0.2864	-0.5115
S KBI-I LIN	-0.4988	-0.3091	-0.4793
S KBI-A LIN	-0.5769	-0.3102	-0.5129
S KBI-N MIN/MAX	-0.6208	-0.2608	-0.4759
S KBI-O MIN/MAX	-0.4988	-0.3091	-0.4793
S KBI-I MIN/MAX	-0.5698	-0.2864	-0.5115
S KBI-A MIN/MAX	-0.6064	-0.3027	-0.5198
S WAI-AVG LIN	-0.7017	-0.5878	-0.6649
S WAI-OP LIN	-0.5169	-0.6214	-0.7368
S WAI-DIN LIN	-0.3558	-0.7012	-0.5913

S WAI-EC LIN	-0.7782	-0.3495	-0.4686
S WAI-AVG MIN/MAX	-0.6715	-0.6011	-0.6711
S WAI-OP MIN/MAX	-0.5715	-0.5595	-0.7179
S WAI-DIN MIN/MAX	-0.3134	-0.7133	-0.5393
S WAI-EC MIN/MAX	-0.7691	-0.3078	-0.4393
ABS WAI-AVG MIN/MAX	-0.5593	-0.6457	-0.6509
ABS WAI-OP MIN/MAX	-0.2514	-0.5932	-0.5539
ABS WAI-DIN MIN/MAX	-0.2536	-0.6768	-0.563
ABS WAI-EC MIN/MAX	-0.7781	-0.3482	-0.4729
ABS SP-AVG MIN/MAX	-0.4669	-0.5556	-0.6682
ABS SP-OP MIN/MAX	-0.1644	-0.4942	-0.5476
ABS SP-DIN MIN/MAX	-0.2538	-0.5845	-0.6052
ABS SP-EC MIN/MAX	-0.6588	-0.324	-0.4986
S SP-AVG MIN/MAX	-0.5137	-0.5387	-0.5545
S SP-OP MIN/MAX	-0.477	-0.5397	-0.6256
S SP-DIN MIN/MAX	-0.2965	-0.6435	-0.4943
S SP-EC MIN/MAX	-0.5868	-0.302	-0.3702

The first exclusion or elimination of indices is based on the final transformation of index scores, either transformed on a minimum/maximum (MIN/MAX) scale to a value between 0-100 or linearly regressed (LIN) to a value between 1-20. The correlations of indices with water quality parameters were highly similar in magnitude across transformation methods (Table 9; 2-9 and 10-17), with no consistent pattern indicating stronger ecological relationships for either scaling approach. Indices transformed on the 1-20 scale were selected based on familiarity in index calculation (IPS, GDI, TDI, BDI, etc.).

The second exclusion of indices is based on the sum product calculation (SP) versus the weighted average calculation (WAI). In the sum product calculation, the index value represents the combined sum of the abundance of species in the community, the optima values for species and the tolerance values (Table 9; 14 – 29). The weighted average calculation takes the sum product value and removes the influence of the abundance and tolerance range after it has been weighted to only present a final optima value for the entire community as weighted by the abundance and tolerance ranges. This calculation shows better correlations with measured environmental parameters and therefore the sum product indices are excluded from the final list. The final exclusion of indices is based on the absolute values versus scaled values (Table 9; 10 – 25).

The indices calculated based on absolute values have better correlations with water quality individually, since these values were not transformed on a scale between 1 and 5.

However, although they have slightly higher correlation with water quality variables, they may only represent the environmental gradient well and not the ecological response of species. Additionally, it is not possible to scale these indices on the 1-20 scale since they are based on absolute values and not on scale values between 1 and 5. Therefore, indices selected for the final presentation are those using scaled optima and tolerance values with a weighted average index calculation and scaled between 1 – 20. The indices using the optima and tolerances inferred from expert knowledge and those calculated through weighted averages and weighted quantiles were selected and the final index scores of these indices are given in Table (A.5). The indices were recoded accordingly: (Knowledge Based Index Average - KBI-A; Knowledge Based Index Organics - KBI-O; Knowledge Based Index Ionic load - KBI-I; Knowledge Based Index Nutrients - KBI-N; Indice de Polluosensibilité Spécifique - IPS; Weighted Average Index Average - WAI-AVG; Weighted Average Index Orthophosphate - WAI-OP; Weighted Average Index Dissolved Inorganic Nitrogen - WAI-DIN and Weighted Average Index Electrical Conductivity - WAI-EC).

In Table 9 the correlations between indices are displayed. The individual indices calculated from optima and tolerance values inferred from expert knowledge correlate well with the average of those indices as weighted according to section, however, the average knowledge-based index correlates better with the IPS index. The IPS index correlates better with the knowledge-based index than the weighted average index, however the weighted average-based index correlates better with measured water quality than the knowledge-based index or the IPS. Similarly, the individual indices calculated using weighted averaging correlate well with the average of those indices according to the same weights and the knowledge inferred index. The weighted average index for DIN does not have strong correlations with any other index, except with the average of the weighted average index and the WA orthophosphate index.

Table 10: Pearson correlations of selected indices (all correlations are significant with $p < 0.05$).

Index	**KBI-A**	KBI-O	KBI-I	KBI-N	**IPS**	WAI-AVG	WAI-OP	WAI-DIN	**WAI-EC**
KBI-A	**-**	0.906	0.970	0.904	**0.820**	0.754	0.660	0.485	**0.724**
KBI-O	**0.906**	-	0.825	0.835	**0.733**	0.759	0.675	0.437	**0.747**
KBI-I	**0.970**	0.825	-	0.783	**0.782**	0.682	0.619	0.513	**0.607**
KBI-N	**0.904**	0.835	0.783	-	**0.773**	0.741	0.597	0.352	**0.801**
IPS	**0.820**	0.733	0.782	0.773	**-**	0.779	0.728	0.532	**0.706**
WAI-AVG	**0.754**	0.759	0.682	0.741	**0.779**	-	0.904	0.761	**0.889**
WAI-OP	**0.660**	0.675	0.619	0.597	**0.728**	0.904	-	0.778	**0.651**
WAI-DIN	**0.485**	0.437	0.513	0.352	**0.532**	0.761	0.778	-	**0.425**
WAI-EC	**0.724**	0.747	0.607	0.801	**0.706**	0.889	0.651	0.425	-

Table 11: Ecological classification of sites as determined by the IPS and other calculated indices in this study.

Index / **Ecological class**	IPS	**WAI-AVG**	WA-EC	WA-DIN	WAI-OP	KBI_A	KBI-N	KBI-O	KBI-I
Eutrophic	84	41	52	5	201	116	234	139	37
Meso-eutrophic	81	131	112	7	78	103	44	139	81
Mesotrophic	63	71	65	29	1	59	2	2	95
Oligo-mesotrophic	18	36	11	56	0	2	0	0	57
Oligotrophic	34	1	40	183	0	0	0	0	10

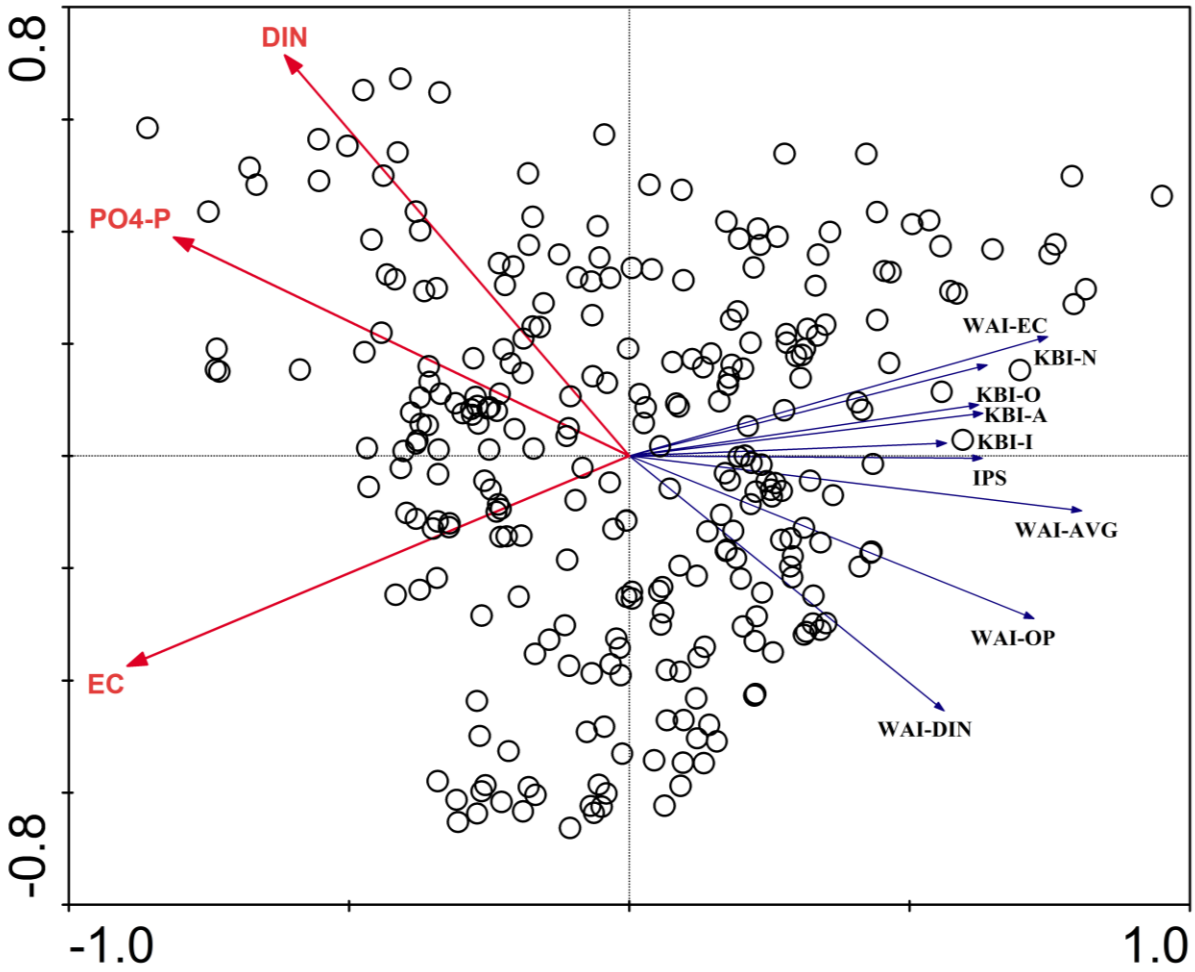


Figure 26: RDA illustrating the relationship between final selected indices and water quality (circles – sites; red arrows – environmental varibales; blue arrows – indices). (DIN – dissolved inorganic nitrogen; EC – electrical conductivity; PO4-P – Orthophosphate).

The RDA illustrates the relationship between the final selected indices and water quality measured. The first axis accounted for 43.7% of the total species variation and 90.9% of the variation explained by the environmental variables. Together, axes 1 and 2 explained 47.5% of the total species variation and 98.7% of the variation attrributed to the environmental variables. The WA indices calculated for EC, DIN and PO_4^{3-}, directly correspond to the measured water quality, as the EC, DIN and PO_4^{3-} increases, the correspnding index values decrease, the WAI-AVG index has a stronger relationship with PO_4^{3-} than EC, however it is still strong (Table 9), and a weaker correlation with DIN. The indices inferred from expert knowledge correlate best with measured EC, as they are all clustered together, indicating they do not show strong relationships with measured DIN

and PO_4^{3-}, however, these indices do show a weak negative correlation (Table 9). The IPS index has a stronger correlation with measured PO_4^{3-} than EC, however these correlations are both strong. The IPS has a weaker relationship with measured DIN.

3.4) Discussion

3.4.1) Optima and tolerances inferred using Weighted Average (WA)

Species optima and tolerances were calculated for EC, DIN and PO_4^{3-} according to the classes defined in tables (5) and (6) respectively. Species are grouped into five trophic preference classes for each variable based on these conversions (oligotrophic, oligo-mesotrophic, mesotrophic, meso- eutrophic and eutrophic), and placed into one of three tolerances classes (specialist, intermediate and generalist) (Appendix A.4). The number of sites used to calculate species optima and tolerances also differ, however all species occur in more than 10 sites, 27 species occur in less than twenty sites, 68 species occur in between 20 and 100 sites and ten species occur in more than 100 sites. Two species occur in more than 200 sites. The occurrence of all species is evenly disturbed across the range of sites, from 10 to 208 sites. The focus of this section will be primarily on the 10 species occurring in more than 100 sites but does encompass trophic preferences and tolerance classes for all species.

For EC, six species fall into the oligohaline group with optima between 100–200 µS/cm. An example of a species categorized as oligohaline for EC, is *Achnanthidium minutissimum* (ADMI – 140 sites). ADMI is placed into the generalist class with a tolerance range of 413 µS/cm (52 - 470 µS/cm) (Figure 18). ADMI is classified as an oligohaline generalist for EC. Twenty-seven taxa occur in the oligo-mesohaline range (200–350 µS/cm). *Gomphonema parvulum* (GPAR – 208 sites) is an example of a species placed into the oligo-mesohaline group. The tolerance range for this species is large (696 µS/cm), occurring in a range from 121 – 817 µS/cm, which explains the occurrence of this species in 208 sites. GPAR is consequently classified as an oligo-mesohaline generalist for EC. The largest number, fifty-one species, are in the mesohaline class with optima between 350–600 µS/cm. Examples of species occurring in mesohaline conditions include *Nitzschia palea* (NPAL – 207 sites), *Nitzschia frustulum* (NIFR – 196 sites), *Eolimna subminuscula* (ESBM – 161 sites), *Cocconeis placentula* (CPLA - 151 sites),

Navicula symmetrica (NSYM – 137 sites), *Amphora pediculus* (APED – 116 sites) and *Mayamaea atomus* var. *permitis* (MAPE – 103 sites). These species have optima calculated for EC between 390 and 520 µS/cm and occur across a wide range of >500µS/cm and are classified as generalists. Nineteen species fall within the meso-euryhaline range (600–1000 µS/cm), however, only one species is classified as meso-euryhaline occurring in more than 100 sites, *Nitzschia veneta* (NVEN – 128 sites). This species has an optimum of 797µS/cm and occurs over a range of 1156µS/cm (419 – 1575 µS/cm). This species is therefore classified as a meso-euryhaline generalist. Only two species have optima above 1000 µS/cm and are classified as euryhaline (*Gomphonema saprophilum* – GPSA and *Amphora veneta* – AVEN). These species occur in less than 20 sites. The optima for these species were calculated using 17 and 14 sites respectively, and so the reliability of these values is low. However, both species are classified as generalist species occurring in a range of (495 – 1600 µS/cm) and (662 – 1491 µS/cm) respectively. Overall, most species are in the mesohaline and oligo-mesohaline ranges, indicating that their EC optima align with moderate to high electrolyte content, with comparatively few species occurring at very high electrolyte levels.

Tolerance values for species for EC fall into three classes (specialist, intermediate and generalist). Of the 105 species, eleven are placed in the specialist class with tolerances below 300 µS/cm. Twenty-one species are placed in the intermediate class, with tolerances between 300 and 600 µS/cm. The remaining seventy-three species are placed in the generalist group, with tolerance ranges above 600 µS/cm. Most species therefore fall into the generalist group, suggesting that while many diatoms have optima in the moderate to high EC range, they still occur across a broad range of electrolyte content.

These optima and tolerances were calculated form datasets collected in the Vaal and Orange Rivers, which are both impacted by salinization through irrigation return flows and agricultural runoff (DWS, 2014; DWS, 2023). Additionally, mining activities around the Vaal River also increased the electrolyte content through effluent runoff (McCarthy *et al*., 2011). These optima and tolerances therefore reflect a community shift adapted for higher electrolyte content and using these optima and tolerances in the EC index calculation reflect salinity changes but may not be sensitive to subtle changes.

For DIN, fifty-six species are placed into the oligotrophic group with optima below 0.5mg/L. Thirty-three species are placed into the oligo-mesotrophic group with optima between 0.5 and 1.5mg/L. Examples of species classified as oligotrophic, with optima between 0.14 and 0.5mg/L, are (*Gomphonema parvulum* – GPAR, *Nitzschia frustulum* - NIFR, *Cocconeis placentula* – CPLA, *Achnanthidium minutissimum* – ADMI and *Amphora pediculus* - APED). Examples of species classified as oligo-mesotrophic with optima between 0.63 and 1.46 mg/L are (*Nitzschia palea* – NPAL, *Eolimna subminiscula* – ESBM, *Navicula symmetrica* – NSYM, *Navicula veneta* – NVEN and *Mayamaea atomus* var. *permitis* - MAPE). Only ADMI is classified as a specialist with a tolerance of 0.43mg/L (0.07 – 0.5mg/L). NVEN, NSYM, ESBM and NPAL are classified as generalists with a range greater than 5mg/L DIN. GPAR, NIFR, CPLA, APED, APED and MAPE are classified as intermediates with a range between 0.5 and 5mg/L. None of the remaining species occur in more than 100 sites. Ten species were placed into the mesotrophic group with optima and tolerance between 1.5 and 2.5mg/L. *Navicula gregaria* (NGRE) and *Achnanthidium (Gogorevia) exiguum* (ADEG) are the only two species occurring in more than 50 sites classified as mesotrophic, with optima of 1.8mg/L and 2.01mg/L respectively. *Navicula gregaria* is classified as a generalist and *Achnanthidium (Gogorevia) exiguum* (ADEG) is classified as an intermediate. Five species are placed into the meso-eutrophic group with calculated optima between 2.5 and 5mg/L. Two species occurring in more than 50 sites are classified as meso-eutrophic, *Sellaphora seminulum* (SSEM) and *Nitzschia amphibia* (NAMP) with optima of 3.49mg/L and 4.01mg/L respectively, both of which are classified as generalists. Only one species is placed into the eutrophic group with an optimum above 5mg/L, *Gomphonema saprophilum* (GPSA – 17 sites) with an optimum of 5.79mg/L, GPSA is also classified as generalist. Overall, most species fall into the oligotrophic group and secondly in the oligo-mesotrophic group, indicating most species do not tolerate elevated DIN.

Twenty-eight species are placed into the specialist class with tolerances below 1mg/L, fifty-nine species are placed into the intermediate group with tolerances between 1 and 5mg/L, and eighteen species are placed into the generalist group with tolerances above 5mg/L. Most species fall in the intermediate class and secondly in the specialist class, indicating most species occur over a narrow range of DIN content. Based on the

calculated optima under DIN, most species do not tolerate elevated DIN and occur over a narrow - to intermediate range of DIN.

The calculated optima and tolerance ranges indicate that these species are associated with low DIN and are not adapted to frequent fluctuations. This aligns with findings from South African rivers, where inorganic nitrogen typically occurs at low background levels (de Villiers *et al*., 2007) and elevated concentrations are usually episodic and locally driven by pollution sources rather than persistent (Tredoux, 2004). As a result, these species are unlikely to be specialised for sustained high DIN conditions

For PO_4^{3-}, none of the species are placed into the oligotrophic group with no optima below 0.005mg/L. Two species are placed in the oligo-mesotrophic group with optima between 0.005 and 0.015mg/L (*Cymbella affinis* – CAFF and *Gomphonema exilissimum* – GEXL). Thirty species are placed into the mesotrophic group with optima between 0.015 and 0.05mg/L, however, only one species occurring in more than 100 sites is classified as mesotrophic - *Achnanthidium minutissimum* (ADMI), with an optima of 0.02mg/L and a tolerance range of 0.012 – 0.056mg/L. Fifty-two species are placed into the meso-eutrophic group with optima between 0.05 and 0.15 mg/L. Species classified as meso-eutrophic for PO_4^{3-} ,and occurring in more than 100 sites, are *Gomphonema parvulum* (GPAR), *Nitzschia palea* (NPAL), *Nitzschia frustulum* (NIFR), *Eolimna subminiscula* (ESBM), *Cocconeis placentula* (CPLA) and *Amphora pediculus* (APED) with optima between 0.05 and 0.15mg/L. Twenty-one species are placed into the eutrophic group with optima above 0.15mg/L, with three species occurring in more than 100 sites - *Navicula symmetrica* (NSYM), *Navicula veneta* (NVEN) and *Mayamaea atomus* var. *permitis* (MAPE), with optima above 0.15mg/L. The former and latter species are classified as generalists for PO_4^{3-}. Examples of specialists for PO_4^{3-} are *Enyonopsis microcephala* (ENCM), *Fragilaria capucina* (FCAP), *Eunotia minor* (EMIN) and *Navicula vandamii* (NVDA), NVDA is classified as a meso-eutrophic specialist, whilst the other species are classified as mesotrophic specialists.

Species' tolerances for PO_4^{3-} are also grouped into the same classes (specialist, intermediate and generalist). Four species are placed in the specialist group, with tolerances below 0.02mg/L. Seventeen species are placed into the intermediate group

with tolerances between 0.02 and 0.1mg/L. Eighty-four species are placed into the generalist group with tolerances above 0.1mg/L. Species' tolerances for PO_4^{3-} indicate that almost all species occur over a wide range of orthophosphate concentration with few species occurring across narrow ranges. Overall optima and tolerances for species for PO_4^{3-} indicate that most specie occur in mesotrophic and meso-eutrophic conditions but occur over a wide range of PO_4^{3}.

The optima and tolerance ranges for species for EC, DIN and PO_4^{3-} indicate most species occur in oligo-mesotrophic and mesotrophic electrolyte content, and in mesotrophic to meso-eutrophic nutrient content, however under EC and PO_4^{3-} most species occur over wide ranges of these variables and for DIN most species occur over narrow ranges. This is supported by frequent shifts in EC and higher electrolyte content in general in the Vaal and Orange rivers, and by the high orthophosphate content originating from agriculture and mining (de Villiers & Thiart, 2007; Mararakanye *et al*., 2022; Erasmus *et al*., 2024).

Most species are classified as generalist species, these species are also found more commonly and therefore, the number of sites used to calculate their optima and tolerances are higher and consequently more reliable. Fewer species are classified as intermediate class species, but the number of sites used to calculate optima and tolerances are still substantial. However, very few species are classified as specialist species. These species occur over narrow ranges of the measured environmental parameters, and therefore their occurrence in sites is also lower, consequently decreasing the reliability of the calculated optima. Therefore, the reliability of calculated optima relies on the occurrence of a species across a wider range of environmental variables. It is important to gather more data to accurately calculate species optima for more rarely occurring species.

The optima and tolerances for species for EC, DIN and PO_4^{3-} were also combined according to the same weights as the index scores (EC – 0.4, DIN – 0.25 and PO_4^{3-} - 0.35). This allows a single optimum and tolerance to be calculated for each species as used in the IPS. None of species occurring in more than 100 sites are classified as oligotrophic and only one species is classified as oligo-mesotrophic – *Achnanthidium minutissimum* (ADMI). *Gomphonema parvulum* is also the only species classified as

mesotrophic occurring in more than 100 sites. Five species are classified as meso-eutrophic (*Niztschia palea* – NPAL, *Nitzschia frustulum* – NIFR, *Eolimna subminiscula* – ESBM, *Cocconeis placentula* – CPLA and *Amphora pediculus* – APED). Three species are classified as eutrophic (*Navicula symmetrica* – NSYM, *Navicula veneta* – NVEN and *Mayamaea atomus* var. *permitis* – MAPE). These species are classified as generalist when combining tolerance values for EC, DIN and PO_4^{3-} (Table 8), except *Achnanthidium minutissimum* (ADMI) which is classified as intermediate. Most species (78) are classified as mesotrophic and meso-eutrophic when combining EC, DIN and PO_4^{3-}, with more than half classified as generalist (66). Only nine species are classified as specialists when combing the variables, these species also occur in less than twenty sites, which confirms their narrow ranges and uncommon occurrence in sites. The remaining thirty species are classified as intermediate.

It's also useful to compare the optima and tolerances calculated here with those used in the IPS. For EC, eight species match the IPS optima and tolerance values exactly, seventeen species have an exact match with the optima (sensitivity) values, and thirty-nine match the tolerance values. For DIN, only one species matches both the optima and tolerance exactly, eight match the optima (sensitivity) value, and twenty-two match the tolerance values. For PO_4^{3-}, fifteen species match both the optima and tolerance values exactly, fifty match the optima (sensitivity) value, and twenty-two match the tolerance values. These matching optima and tolerance values likely contribute to the strong correlations between the IPS score and each of the individual indices. PO_4^{3-} has the most matches overall, followed by EC and then DIN, and this trend is reflected in the Pearson correlation coefficients of 0.74, 0.68 and 0.52 for PO_4^{3-}, EC and DIN respectively. The correlation with the IPS and the WA-OP index is supported through the original purpose of the IPS, which was designed to focus on long-term trophic changes and not short-term nutrient spikes (Coste in CEMAGREF, 1982).

These optima and tolerances therefore support the applicability of the indices calculated, where each of the individual indices, WA-EC, WA- PO_4^{3-} and WA-DIN can be used to indicate salinization through EC changes, nutrient enrichment through increased PO_4^{3-} and changes in DIN respectively, and the combined average index can be used to assess overall water quality. These optima and tolerances correlate with those found in the IPS

and therefore support the use of WA to calculate optima and tolerances for index calculation.

3.4.2) Optima and tolerances inferred using Generalised Logit Regression (GLR)

Using GLR was explored as a method of calculating species optima and tolerances. However, this method was ultimately rejected due to statistical and ecological limitations of the method. GLR estimates species optima using a quadratic curve (e.g. gaussian curve). To accurately fit such a curve for species optima requires the use of large datasets where species are evenly distributed along the environmental gradient (ter Braak & Looman, 1986). Additionally, sparse and zero inflated datasets are not ideal for optima calculation using GLR (Zuur *et al*., 2009). The dataset in used in this study contains species abundances unevenly spread along the environmental gradient with species occurring in clusters along the gradient. This makes using GLR to estimate optima difficult, additionally, the consistent occurrence of species along the gradient is also low, with some species occurring in patches along the gradient instead of clustering in the extreme ends of middle of the environmental gradient. Only a small number of species occurred frequently enough along the environmental gradient to calculate their optima using GLR, however, using WA on these species already provide an accurate estimate of calculated optima. GLR is sensitive to rare species and only works well when species are abundant and evenly spread along the environmental gradient (ter Braak & Looman, 1986; ter Braak, 1987; Zuur *et al*., 2009). Ecological datasets are often sparsely populated, and species abundances are often zero inflated which makes using GLR very difficult. WA is recommended for ecologically sparse datasets that are populated by rare species and clustered distribution and therefore this method was chosen instead of GLR (ter Braak & Looman, 1986; ter Braak, 1987).

3.4.3) Calculated indices

According to the weighted average formula used to calculate index scores, the abundance-sensitivity value (ASV) is divided by the abundance value (AV), meaning both optimum and tolerance influence the index (formula 4 section 3.2.2). A high optimum (greater than 3) increases the index score because it contributes more to the numerator,

whereas a low optimum (2 or less) contributes less and reduces the score. The tolerance further modulates this effect: for a species with a high optimum, a higher tolerance (specialist) amplifies the positive influence on the index, while for a species with a low optimum, a higher tolerance amplifies the negative influence. In contrast, generalist species (low tolerance) have a muted effect on the score regardless of their optimum. The differences between species optima and tolerances as calculated using WA, knowledge inferred and recorded in the Indice de Polluosensibilité Spécifique (IPS) are responsible for the difference in index scores. Additionally, the presence or absence of species in the IPS also contributes to a difference in the index scores.

The indices are used to classify sites according to Table (7) in trophic classes from oligo-trophic to eutrophic. Table (11) displays the ecological classification of sites for each index calculated, site classification for each dataset (1-5) is given in appendices (A.5 – A.9). The IPS classifies eighty-four sites as eutrophic, eighty-one sites as meso-eutrophic, sixty-three sites as mesotrophic, eighteen sites as oligo-mesotrophic and thirty-four sites as oligotrophic. Most sites are classified as eutrophic and meso-eutrophic according to the IPS and, which reflects the water quality and trophic conditions of the Vaal and Orange rivers. Few sites are classified as oligo-mesotrophic.

The following species are missing from the IPS dataset and therefore they do not contribute to the final index score (*Encyonopsis leei* var. *sinensis* (ENLS), *Planothidium engelbrechtii* (PENG), *Cymbella kappii* (CKPP), *Gomphonema latistigmata* (GLST), *Gomphonema venusta* (GVNU), *Navigiolum adamantiforme* (NADF) and *Navicula nielsfogedii* (NSLT)). These species are mostly recorded in dataset 3 and 4. However, only CKPP and NADF are recorded in dataset one and PENG is mostly recorded in dataset 5. The occurrence of these species in the datasets contributes to the difference in sites index scores, especially where these species are found in high abundance. Therefore, the difference in site scores not only stems from the difference in optima and tolerances, but the occurrence of species not included in the IPS.

Dataset 1 is the largest dataset with 117 sites. In dataset 1, the IPS classifies most sites as meso-eutrophic (44), mesotrophic (34) and eutrophic (34). Only five sites are classified as oligo-mesotrophic and none classified as oligotrophic. The WAI-AVG index also

classifies most sites as meso-eutrophic (55), and the remaining sites as either mesotrophic (47) or eutrophic (14). Only one site is classified as oligo-mesotrophic. Therefore, the IPS and WAI-AVG indices both classify most sites as meso-eutrophic, however, the WAI-AVG index classifies fewer sites as eutrophic than the IPS and more sites as mesotrophic. This is due to the occurrence of CKPP and NADF in dataset 1, which increases the index scored calculated using WA. The individual indices indicate these sites have moderate to high electrolyte content, with low DIN and high orthophosphate concentrations, which is reflected by the WQ in this dataset (Table A.5). The KBI-A index classifies most sites as eutrophic and meso-eutrophic, only nine sites are classified as mesotrophic and no site is classified as either oligo-mesotrophic or oligotrophic. The KBI-N and KBI-O indices contribute most to this classification indicating these sites are high in nutrients and organic matter. The KBI-I index classifies more sites as mesotrophic than the KBI-N and KBI-O indicating these sites have moderate to high electrolyte content. Overall the IPS, WAI-AVG and the KBI-All indices indicate these sites are classified as mesotrophic to eutrophic, having high nutrient content and moderate to high electrolyte content.

Dataset 2 contains 49 sites, which are mostly classified as oligotrophic by the IPS. However, sites in this dataset are also classified as eutrophic by the IPS, with few sites classified as oligo-mesotrophic to meso-eutrophic. The WAI-AVG index follows the same trend, however, more sites are classified as meso-oligotrophic than oligotrophic. This is due to the influence of the WAI-OP index decreasing the average index score. The WAI-EC index however, follows the classification of the IPS almost exactly. Similarly, the WA-DIN index classifies most sites as oligotrophic and other sites as eutrophic with few sites classified as oligo-mesotrophic to meso-eutrophic. Therefore, the WAI-AVG index is mostly determined by the WA-EC and WA-DIN indices but downweighed by the WA-OP index. The WA-EC index correlates strongly with the IPS. The KBI-A index however, classifies most sites are meso-eutrophic and eutrophic, which is caused by the KBI-O and KBI-N indices. The water quality reflects the accuracy of the IPS and WA indices, since most sites have PO_4^{3-} concentrations below 0.015 mg/L, low DIN and low electrolyte content. The remaining sites have high PO_4^{3-} concentrations (> 0.03mg/L) and high electrolyte content (> 1000 µS/cm) which is reflected by the index scores. The sites with

the highest EC, PO_4^{3-} and DIN are sites D2-9, D2-29, D2-30 and D2-31 (Table A.4), which are sites with the lowest index score (highest water quality impacts) according to all the indices calculated. The KBI indices less accurately reflect the water quality at these sites than the WA indices and the IPS. The IPS has optima and tolerances for all species recorded in this dataset, thereby substantiating the high correlation between the IPS and the WA indices.

Dataset 3 has 60 sites. These sites are mostly classified as eutrophic by the IPS with fewer sites classified as meso-eutrophic and mesotrophic. The WAI-AVG index classifies most sites as meso-eutrophic and eutrophic. With few sites classified as mesotrophic to oligo-mesotrophic. This classification is mainly determined by the WAI-EC index which classified most sites as meso-euryhaline but no sites as euryhaline, the sites classified as eutrophic by the WAI-AVG are determined by the WAI-OP index which classified more than half of the sites as eutrophic. The WAI-EC index indicates these sites have moderate electrolyte content while the WAI-OP indices indicate these sites have high PO_4^{3-} concentrations, together these indices reflect the correlation of the WA-AVG index with the IPS. The IPS indicates these sites have moderate to high electrolyte content and high nutrient loads. This is reflected by the water quality with most sites having high PO_4^{3-} (> 0.02mg/L) and more than half of sites having moderate to high electrolyte content (>350 µS/cm). The KBI-A index classifies most sites as eutrophic and meso-eutrophic, which is equally determined by the KBI-N and KBI-O indices. The KBI-I index slightly increases the KBI-A index score, with more sites classified as mesohaline based on electrolyte content. Considering the water quality of sites in this dataset, the IPS, WA indices and KBI indices accurately reflect the water quality, however the KBI indices more accurately reflects the water quality in terms of high nutrient content and moderate electrolyte content. This dataset also contains most species the IPS does not include in its calculations. However, the scores do not differ substantially across sites. Indicating that although some species are not included in the IPS, those that are included have a higher relative abundance than those not included.

Dataset 4 contains 39 sites. Most sites are classified as eutrophic and mesotrophic by the IPS, some sites are also classified as meso-eutrophic and few sites are classified as meso-oligotrophic and oligotrophic. The WAI-Avg index classifies most sites as meso-

eutrophic with few sites classified in other trophic categories. This classification is mostly determined by the WA-EC and WA-OP indices, where the WA-EC indicates moderate to high electrolyte content and the WA-OP indicating high nutrient content. Similarly, the KBI-A index also classifies most sites as meso-eutrophic but also as eutrophic. Which is mainly due to the influence of the KBI-N and the KBI-O indices. The sites in this dataset as classified by the WA and KBI indices have moderate electrolyte content and high nutrient content. The IPS classified most sites as having moderate to high electrolyte content and moderate to high nutrient content. The water quality reflects these sites have moderate electrolyte content (350 - 600µS/cm) and high PO_4^{3-} content (>0.03mg/L). Therefore, the individual indices calculated using WA and inferred from expert knowledge, more accurately reflect the water quality of these sites than the IPS and the average of the WA and KBI indices.

Dataset 5 has only fifteen sites. Most of the sites are classified as meso-eutrophic and eutrophic by the IPS. The WAI-AVG index reflects the same classification but more sites are classified as eutrophic rather than meso-eutrophic. The KBI-AVG reflects the same classification as the WAI-AVG index. The correspondence of these indices is reflected by the WAI-EC and KBI-I indices indicating sites have moderate to high electrolyte content, and the WAI-OP, KBI-N and KBI-O indices indicating high nutrient content in these sites. The WA and KBI indices as well as the IPS indices accurately reflect the water quality of these sites, with most sites having high PO_4^{3-} (>0.04 mg/L) and moderate to very high electrolyte content (350 – 1600 µS/cm).

Overall, the WAI, KBI and IPS indices all accurately reflect site water quality within all datasets (Figure 26). Although some optima and tolerances for species are not included in the IPS, the entire community structure at sites where these species are not present accurately reflect the water quality as calculated by the IPS. Therefore, diatom assemblages are robust and accurately reflect the same water quality conditions, even if the community structure is composed of different species with similar optima and tolerances. Furthermore, the abundance of common species is naturally higher than rare species and the optima and tolerances calculated for those species can be more reliably used in the index calculation. And because these species occur in greater abundance, the calculated index scores are accurate.

The RDA shows how the individual indices indicate specific water quality, where the IPS, WAI-AVG and the inferred knowledge indices might reflect average water quality. Therefore, in sites where the WA-AVG, inferred knowledge indices and IPS reflect average water quality the individual indices can provide insight as to why a particular score is obtained. The use of the individual and combined indices in the context of water quality monitoring in South Africa offers a robust analysis of trophic state and nutrient loads in south African rivers than can potentially replace the IPS as a preferred index of calculation. However, more data is needed to accurately calculate species optima and tolerances before the IPS can be replaced. Therefore, these newly proposed indices should be used in tandem with the IPS during biomonitoring, however, the individual indices may provide better insight into specific water quality degradation, either nutrient loads, organic pollution or salinization. The IPS can only provide an average of these conditions and does not indicate any specific type of pollution. Weighted Averaging proved an effective method for accurately calculating species optima and tolerance using a small dataset. Similarly, inferred knowledge proved just as valuable to the creation of diatom indices.

In combination the WA method can be bolstered by inferred knowledge and the use of larger datasets to more accurately calculate species optima and tolerances. Several of the datasets used in this study are sampled form the Vaal and Orange rivers, which form part of the Vaal-Orange basin, the largest in South Africa. This basin provides water for agriculture, industry and domestic use and is one of the largest contributors to freshwater for these practices. Consequently, these rivers are highly degraded. The diatoms present in these rivers are adapted to these types of pollution and degraded conditions. Therefore, although optima and tolerances for only 105 species were calculated and used in this dataset, they represent a large part of the freshwater diatom flora found in the central region of the country.

Therefore, the application of WA and inferred knowledge to calculate optima and tolerance for diatom index creation in South Africa was successful and the resulting indices accurately reflect water quality and are on par with the IPS index. Whilst in addition providing more specific information on salination, nutrient loads and organic contamination than does using the IPS alone.

3.5) Summary

Novel diatom indices for water quality monitoring in South Africa must accurately reflect the most common and pressing water quality issues in the country: eutrophication, salinization and organic pollution. Diatoms form an integral part of aquatic food webs and respond reliably and predictably to a wide range of pollutants and environmental changes. Although diatoms already play a key role in the South African biomonitoring framework, no region-specific indices currently exist.

The IPS index is the most widely applied diatom index in South Africa and has been largely successful in assessing river condition in terms of nutrient enrichment, organic loading and ionic composition (a combined measure). Developing new regional diatom indices to reflect water conditions was based on first calculating species optima and tolerance ranges using the weighted average approach applied in the IPS. Datasets from the Orange–Vaal basin amongst others were used to infer optima and tolerances for 105 species using weighted averaging and inferred knowledge, and these were incorporated into a comparable weighted-average calculation.

Separate indices were calculated for each environmental parameter representing nutrient content, organic load and ionic composition, and these were combined into single multimetric indices. The newly developed indices correlate well with measured water quality in a similar way to the IPS. However, the individual component indices more accurately reflect the specific pollution types contributing to water quality deterioration and can be used as stand-alone reporting metrics. Therefore, these new indices can reliably be used to infer water quality degradation in South African rivers and should be used alongside the IPS until larger datasets become available to refine species optima and tolerance estimates with greater accuracy.

CHAPTER 4: ACID MINE DRAINAGE (AMD) INDEX DEVELOPMENT

4.1) Introduction

Wetlands, as defined by the South African National Water Act (NWA, 1998), are transitional areas between land and water where the water table is near the surface or the land is periodically submerged, supporting vegetation adapted to saturated soils. In South Africa, wetlands are classified as aquatic systems due to the periodic presence of water (Ollis *et al.*, 2006). It is crucial to maintain the natural health and function of wetland ecosystems as they provide habitat for many endemic species and contribute to the rich diversity of animal and plant life in South Africa, in addition, wetlands provide many ecosystem services to humans of which flood control and water purification is of particular importance (Collins, 2005).

However, despite their importance, a substantial percentage of South African wetlands are threatened (79%), and 61% are not protected (van Deventer & Nel, 2025). Many wetlands are subject to anthropogenic impacts that include industrial pollution and acid mine drainage (AMD). AMD is among the multitude of impacts that threaten wetlands surrounded by coal deposits. Coal mining is a major contributor to the economy of the country. Domestically, coal mining is important for electricity generation, fuel production and industrial use. Seventy-two percent of the country's energy demands are satisfied through coal mining (Ratshomo & Nembahe, 2016). The Mpumalanga Highveld's coal-rich region, especially the Highveld coalfield, significantly supports the local economy (Fosso-Kankeu *et al.*, 2016).

AMD is a consequence of coal mining, as coal is extracted from the earth, pyrite is extracted along with it. Pyrite contains sulphur-rich minerals and when these minerals encounter water, they dissolve and acidify the effluent and runoff from coal mines (Ochieng *et al.*, 2010). As wetlands are natural depressions and accumulate water, salts and nutrients, acidic runoff also flows into wetlands and may impact the biota within. AMD is characterised by high concentrations of sulphate, which in turn elevates electrical conductivity within wetlands and lowers the pH (Smucker & Vis, 2009). Wetlands that have a natural assimilative capacity and buffering ability will withstand much of the effects

of AMD, however, when AMD becomes frequent and more concentrated, the buffering capacity is exceeded, and detrimental effects become apparent (Humphries *et al*., 2017).

Biomonitoring has long been implemented in South Africa to detect environmental changes within aquatic ecosystems. In South Africa, wetland health is gauged through physical, functional, and biotic indicators (Kotze *et al*., 2012), with biological assessments quantifying deviations from natural states. These often use species composition and abundance to reflect disturbance. Aquatic organisms such as diatoms, macrophytes, macroinvertebrates, and fish are well-established bioindicators in river systems (Kleynhans, 1999; Dickens & Graham, 2002; Taylor *et al*., 2007a; Dallas, 2007). Diatoms are globally recognized bioindicators thanks to their rapid reproduction, environmental sensitivity, and broad habitat presence (Hattikudur *et al*., 2015; Lobo *et al*., 2016). Their ecological preferences for factors like pH, electrical conductivity, organic content, and nutrient levels are well-documented (Lobo *et al*., 2016). In AMD-affected areas, certain diatoms outperform other indicators due to higher acid tolerance (Battarbee *et al*., 2001; De Nicola, 2000). Their quick generational turnover (approximately two weeks) enhances responsiveness to water quality shifts (Harding *et al*., 2005), and they are effective indicators of organic pollution, eutrophication, salinity, acidification, and heavy metal contamination (Wu, 1999; Tapia, 2008).

A study by Riato *et al*. (2018) developed a diatom-based multimetric index tailored to AMD-impacted wetlands. These indices, specific to wetland types, used multiple metrics to capture various structural and functional aspects of diatom assemblages. Metrics fell into three categories: similarity to reference sites, functional group composition, and taxonomic diversity. Typology-based classification assisted in accounting for natural variation. This was further tested and applied over several seasons by Erasmus (2021).

The structure of diatom communities, particularly their growth-forms, life-forms and ecological guilds, provide insight into the ecological condition of aquatic habitats. Growth-form, as defined by Hudon & Legendre (1987), reflects adaptations for life strategies and resource competition, characterised by form (solitary or colonial), posture (prostrate or erect), and mobility. Diatoms are grouped into life-forms based on different life stages: solitary, motile, attached and colony-forming, and grouped into ecological guilds based

on resource use strategies: low-profile, high-profile, motile, and planktonic (Rimet & Bouchez, 2012b).

Diatom communities shift with environmental change, partly due to their colonization ability. Colonization begins with adnate and motile pioneer species, forming biofilms that support later-stage erect forms, which dominate climax communities (Totti *et al.*, 2009). Growth-forms, life-forms and guilds offer practical alternatives, responding predictably to stressors like nutrients and organic pollution (Passy, 2007; Rimet & Bouchez, 2011, 2012a).

The index developed by Riato (2017) is effective specifically for depressional wetlands. The present study features a variety of wetlands including seeps, floodplain wetlands as well as channelled and unchanneled valley-bottom wetlands which may also be affected by AMD. The application of a diatom-based indices developed for South African rivers has potential for use in wetlands but are not as effective as indices developed explicitly for wetlands. Likewise, the application of AMD multimetric indices developed for depressional wetlands in other wetland types is limited as these are different hydrogeomorphic units.

The aim of this chapter is to develop a diatom-based index for application in wetlands impacted by AMD. The study aims to determine the diatom communities associated with different wetlands to develop an index that can distinguish between AMD-impacted and non-impacted wetlands. The diatom-based index has potential for integration into existing South African monitoring frameworks for wetland assessment by providing a diatom-based approach specific to wetlands affected by AMD. The index could be valuable to the Department of Water and Sanitation (DWS), the South African National Biodiversity Institute (SANBI) and provincial management authorities that assess wetlands impacted by mining activities. Additionally, the index is supported by an executable software tool that can enhance the capacity of routine monitoring, data capturing and rehabilitation assessment of wetlands.

4.2) Methodology

A dataset containing data from forty-two sites with corresponding diatom abundances and measured environmental parameters was used to create the multimetric index. Site selection was done to include reference sites (reference = D-1) as well as disturbed sites (D-2, D-3 and D-4), disturbance was determined by site proximity to mining activity and confirmed with water quality, specifically measured sulphate concentrations (Table B.3), the closer the site to mining activity and the higher the sulphate concentration, the greater the allocated disturbance level. Sites with corresponding wetland type and disturbance levels are displayed in Figures 27- 29. Each site was sampled twice for diatoms and water quality, once during the wet season (summer rainfall) and once during the dry season (winter), however, two sites were not resampled during the wet season. Therefore, of the 45 sites, only 43 were resampled. A total of 88 sites were sampled across both seasons. Additionally, two separate geographical areas were selected, half of the sites were sampled in the eMalahleni area in Mpumalanga and the other half in the New Castle area in KwaZulu-Natal.

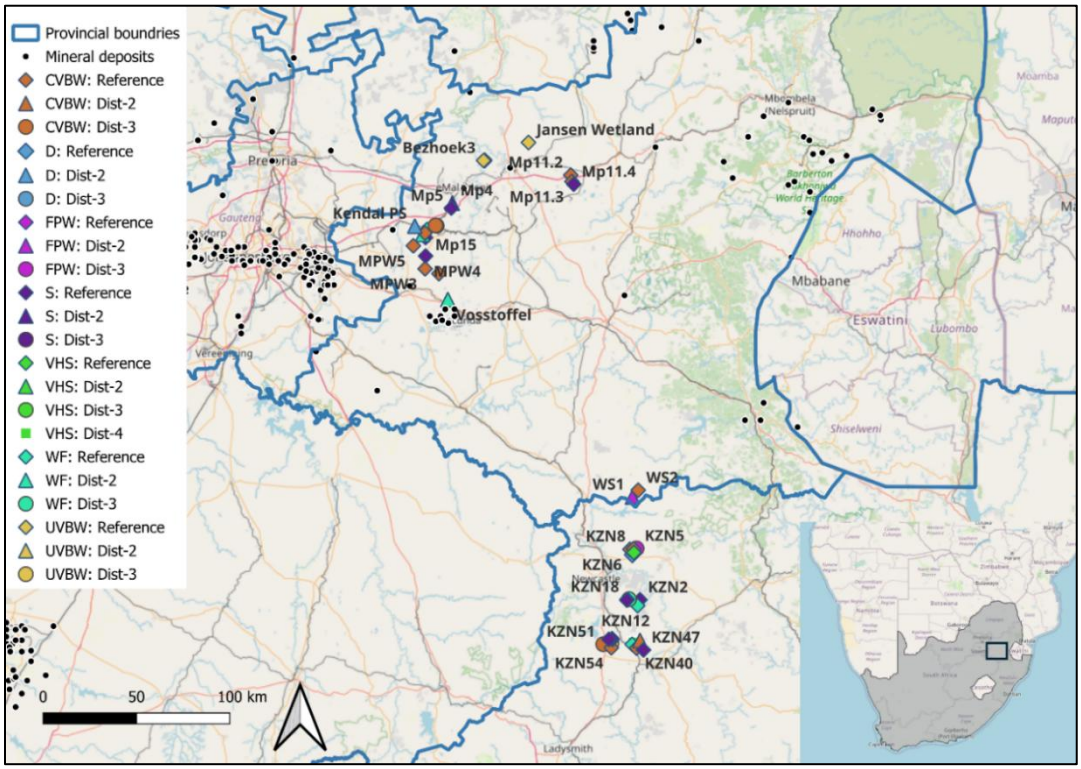


Figure 27: Map of sampling sites in Mpumalanga (top) and KwaZulu-Natal (bottom). Sites are coloured based on wetland type (Table B.4). Symbols represent disturbance levels from 1 – 4 during the dry season.

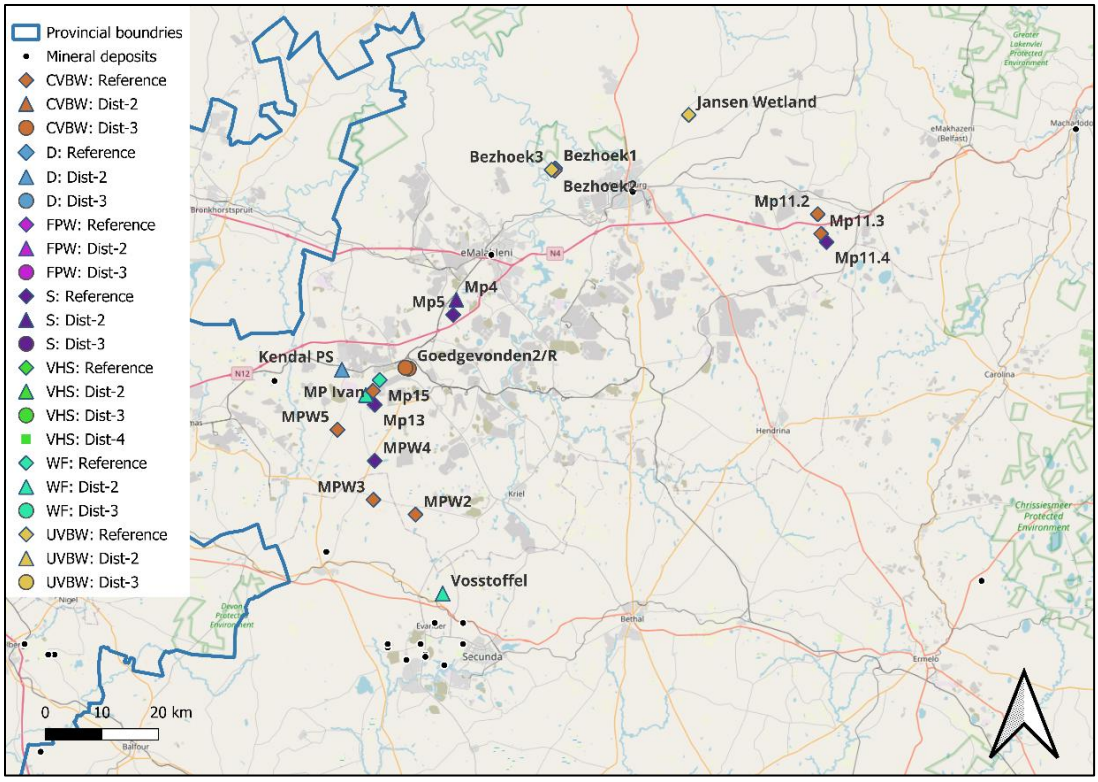


Figure 28: Map of sampling sites in the Mpumalanga province. Grey areas on the map indicate mining activity.

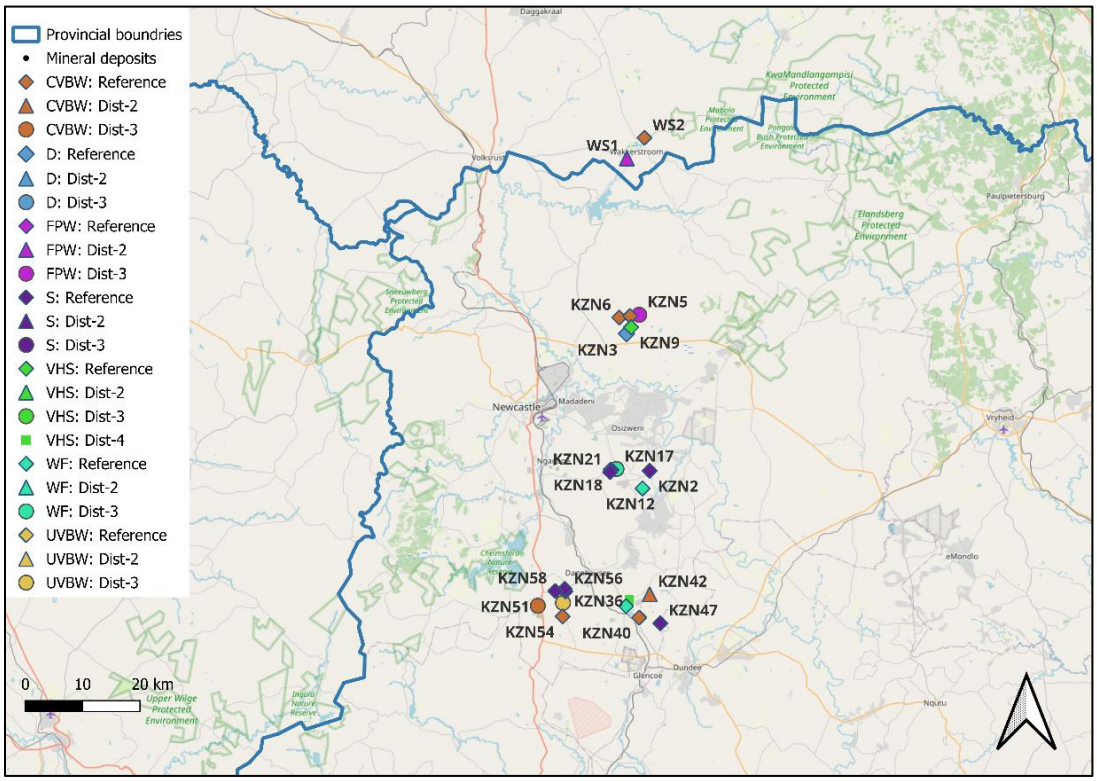


Figure 29: Map of sampling sites in the KwaZulu-Natal province. Grey areas on the map indicate mining activity.

The data obtained therefore integrates geographical and seasonal variation in diatom communities as well as variation in the communities because of AMD disturbance by sampling reference and disturbed sites. Reference and disturbed sites were determined based on the proximity of a site to an active coal mine. Reference sites were situated further from coal mines than disturbed sites, without impacts from mining. Epiphytic and epilithic diatoms were sampled and processed according to Taylor (2007b) with acid digestion using the hot HCl/H_2O_2 method. Samples were cleaned and dried on coverslips before permanent mounting with Pleurax (Von Stosch, 1974). At least 400 diatom frustules were counted for each site and taxa were identified to species level using a Nikon 80i microscope under 1000x magnification with Differential Interference Contrast (DIC).

Before calculating the MMI index scores, the optimal environmental conditions (μ) for taxa were calculated. Life form, AMD tolerance and osmotic tolerance, were also ascribed numerical values as these are included as weights in the index calculation. The criterion used to include a taxon in the calculation of optimum environmental conditions was that the taxon must appear in a minimum of 10 sites with more than 1% abundance. Of the 290 taxa identified a total of 78 (26.9%) taxa satisfied this criterion and a μ-value for each was calculated through weighted averaging (WA). The purpose of weighted averaging was to determine an average value of a variable with the addition of assigning weights to metrics that influence the variable. Therefore, the average obtained is influenced by the weight given to a certain parameter. In the case of determining the optimal environmental condition of diatoms, the relative abundance of a taxon serves as the weight for calculating the average value of an environmental variable under which the diatom is likely to be found under high abundance and high fitness.

The optimum environmental condition (μ) calculation is based on the following formula:

$$\mu = \sum_{i=1}^{n} Y_{ik} X_i / y + k \ , \ y + k = \sum_{i=1}^{n} Y_{ik} \qquad (\mathbf{1})$$

Where $\boldsymbol{Y_{ik}}$ is the abundance of the k-th taxon in the I-th region and $\boldsymbol{X_i}$ is the environmental variable in the I-th region (Shin *et al.*, 2022).

Using formula (1), a weighted average optimum was calculated for each diatom taxon based on the concentration of the environmental parameter under which they reach the highest relative abundance in the community.

Diatoms fill functional niches within their community and either facilitate or hinder the growth of other diatom species in the community. Diatom species also have physiological adaptation to thrive under varying degrees of nutrient concentrations, organic pollution, and light availability (Patrick, 1977). However, this is largely accounted for through calculation of the optimum environmental condition. Diatoms also have different life forms and growth forms and have certain tolerances toward specific types of pollution, including AMD disturbance, and changes in osmotic pressure (Patrick, 1977; Parikh *et al*., 2025).

Therefore, these features were also considered as metric in the calculation of the multimetric index. From literature (Riato *et al*., 2018; Erasmus, 2021) and prior knowledge, it was determined which diatoms included in this study are considered tolerant to AMD disturbance and osmotic fluctuations and which are not and they were then assigned a value accordingly; diatoms sensitive to osmotic change and AMD impact receive a value of two, whilst those that are tolerant receive a value of one; a diatom sensitive to osmotic change and AMD disturbance receives a higher score than a diatom that is tolerant to these changes. Concerning life-forms, an attached life-form received the highest score. This is because attached diatoms cannot move/or move more than locally to avoid any disturbance, source of pollution or nutrient contamination (Rimet & Bouchez, 2012b). A scale value was assigned based on the growth form of diatoms; attached diatoms were assigned a value of three. However, not all attached taxa tolerant to these changes respond in a similar way, therefore it is important to include the optimal environmental conditions for each taxon in the calculation.

Tube-forming diatoms were assigned a value of two. Tube-forming taxa as a group construct a muco-polysaccharide tube to live in, this tube serves as a shelter for protection but also allows these taxa to occupy greater parts of a 3-D environment by growing out of the community toward the surface of the biofilm (e.g. *Frustulia crassinervia*). Conditions under which diatoms construct tubes or stalks to occupy a greater 3-D space are usually correlated with a competition for nutrients and light, under such conditions the biomass

of the community is high and primary production is also high which is a result of eutrophication. However, in a wetlands system, tube-dwelling diatoms build tubes for protection against osmotic fluctuations and to create a more favourable microenvironment inside the tube (Hormiga *et al*., 2025).

Motile diatoms are assigned a value of one. These diatoms may move away from disturbance if they are not adapted to changing conditions and move away if nutrient concentrations are unfavourable (Rimet & Bouchez, 2012b). Therefore, these motile taxa often avoid disturbance through their movement, however, if such a motile taxon does not move away from disturbance or contamination it is reasonable to assume these conditions are tolerable. Therefore, motile diatoms receive a low score. Motile diatoms are also considered tolerant to higher sediment loads, as being motile they can move through the sediment and towards the light to photosynthesise. Therefore, the combinations of these metrics in an index calculation would produce a higher score if a site were free from AMD disturbance and a lower score if a site is disturbed.

The calculation of the multimetric index value consists of three steps. The first step is to calculate a standalone index value for each of the five parameters using diatom optimum conditions, life forms, AMD tolerance, osmotic tolerance and relative abundance as metrics (formula 2), as calculated above. Thereafter the calculated index values are weighted based on the relative importance of the environmental variable to AMD and summed to calculate a final MMI value (formula 3). Finally, the MMI values are normalised across sites for comparison and statistical analysis.

A site index value for each of the five parameters is determined by using the obtained optimum environmental condition of each diatom species in the community (formula 1), weighted by its abundance, life form, osmotic tolerance and AMD tolerance, and has the following form:

$$S_{(paramter\ j)} = \sum_{i=1}^{n}(\mu_i \times a_i \times o_i \times lf_i \times z_i) \qquad (\mathbf{2})$$

Where $\boldsymbol{S}$ is the community score for a given parameter *j*, $\boldsymbol{\mu_i}$ represents the optimal environmental condition for the i-th species, $\boldsymbol{a_i}$ represents the abundance of the i-th species, $\boldsymbol{o_i}$ represents the osmotic tolerance of the i-th species, $\boldsymbol{lf_i}$ represents the life form of the i-th species and $\mathbf{z_i}$ represents the AMD tolerance of the i-th species. This formula

produces an index score for each of the five parameters by weighting the optimum environmental condition of each species in the community with its respective abundance, osmotic tolerance, life form and AMD tolerance. This was repeated for each environmental parameter amounting to five index scores per site, one for each of the environmental parameters chosen.

The second step is to combine these five index scores into one score by assigning weights to each environmental variable (formula 3), because they have varying levels of importance when associated with AMD disturbance. Sulphate receives the highest score since it is a principal factor that influence and correlates with AMD contamination in South Africa (Humphries *et al*, 2017). Electrical conductivity and pH receive intermediate scores as they indicate acid generation and ionic enrichment but are not necessarily linked to sulphate concentrations specifically (Leiva *et al*., 2021). Alkalinity and chloride were assigned low weights as they are less diagnostic of AMD but do indicate on buffer capacity and ionic balance (Lee & Kim, 2008). Therefore, weights are assigned to environmental parameters based on their importance to AMD disturbance and were guided by established literature and expert opinion. Weights were assigned accordingly; Sulphate – 0.3, pH – 0.2, EC – 0.25, chloride – 0.1, alkalinity – 0.15.

The cumulative value of the weights for each parameter amounts to one, ensuring relative contributions of each parameter to the site score and ensuring the scale remains the same. By using these weights, we can combine the five scores into one score by using the following formula:

$$Index\ (S) = \left(\boldsymbol{S}_{sulp} \times\ W_{sulp}\right) + \left(\boldsymbol{S}_{pH} \times\ W_{pH}\right) +\ \left(\boldsymbol{S}_{EC} \times\ W_{EC}\right) +\ \left(\boldsymbol{S}_{chl} \times\ W_{chl}\right) +\ \left(\boldsymbol{S}_{alk} \times\ W_{alk}\right) \quad (\mathbf{3})$$

Where ***S*** represents the score for each environmental parameter and ***W*** represents the corresponding weight of each parameter. This index score represents the sum of the weighted average score for each environmental parameter. The index score is calculated for each site.

Once the index scores for each environmental parameter were weighed, they were normalized between 0 and 100 before summation, it was considered whether to normalize between the 5th and 95th percentiles but this results in some scores being negative.

Normalization was done separately for each environmental parameter using the global minimum and maximum scores of that parameter. After normalization, the scale for sulphate and EC index scores were reversed using the formula: (100 – value). This reversal is a key step in the calculation in the final index, the index values for pH, alkalinity and chloride increase with higher concentrations of those parameters, which is a positive sign for wetlands systems not impacted by AMD. On the other hand, the index values for sulphate and EC increase under increasing sulphate and EC concentrations which is a sign of AMD impact in wetlands, therefore the scale is reversed so high values of the sulphate and EC index score correlate with decreasing sulphate and EC concentrations. Therefore, when combining these scores in the above formula an increase in the index value correlated with less AMD disturbance relative to other sites.

The third step is to normalise and compare index scores across all sites. The index scores do not possess any inherent meaning; it is only when comparing all index scores to one another that and indication is given of which sites are impacted and which are not. To compare these scores, they must again be normalized to fit the same scale. The Index values are not normally distributed and so using z-score normalization is not feasible, normalization using minimum and maximum values was again undertaken:

$$MMI = Index_S - Index_{min} / Index_{max} - Index_{min} \quad (\mathbf{4})$$

Where the MMI value is the index value of a site subtracted by the minimum value across all sites and divided by the difference between the Index maximum and minimum across all sites.

CCA and RDA was used to graphically illustrate the ordination of species data, sites sampled, environmental parameters measured and calculated index scores. CCA was used to illustrate the temporal variation in species composition at sites sampled before calculating optimal environmental conditions as well as to ordinate the observed disturbance levels at sites sampled with measured environmental parameters measured. RDA was used to ordinate calculated index values with measured environmental variables, additionally, observed disturbance sites were independently added to the ordination and illustrated with convex hulls.

Furthermore, to assess the reliability of the calculated index scores, bootstrapping was performed in the absence of additional datasets. One thousand iterations of the index calculation were carried out by resampling the optimal environmental conditions for each parameter, with each value adjusted by 2% of its respective range. The resulting index scores were then validated using Pearson correlations to confirm the consistency of the index.

The index calculation process has been streamlined through the development of a user-friendly executable software program. The executable program was developed in Python and complied as a standalone windows application. Users input species and environmental data and perform the index calculation. The program automatically matches the provided species names to a standardized list containing the optimal environmental conditions, AMD tolerance, osmotic tolerance, and life-form for each taxon. Once matched, the program performs the necessary calculations and generates an Excel spreadsheet listing each site from the input file along with its calculated index score for easy comparison and interpretation.

4.3) Results

CCA was used to determine any spatial and temporal variation in diatom community structures across sites as well as to determine how much variation in the diatom community structure is due to the five environmental parameters selected (pH, sulphate, EC, alkalinity and chloride). RDA analysis was used to illustrate the relationship between final index scores, environmental parameters and sites, with observed disturbance levels used to group sites. Correlation analysis was also used to determine the correlation strength between final index scores and environmental parameters.

4.3.1) Canonical Correspondence Analysis (CCA)

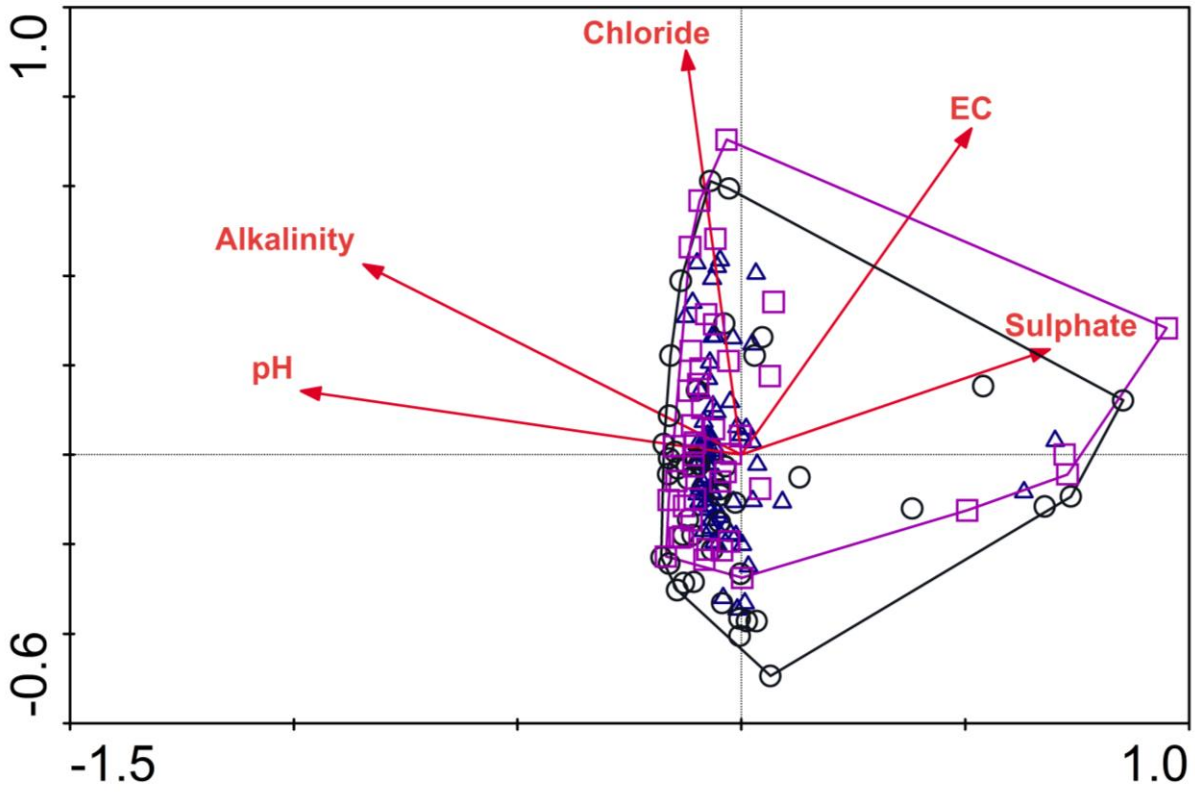


Figure 30: CCA plot illustrating the relationship between environmental variables and species composition at sites. Convex hulls represent season, Black circles – Dry season; Purple squares – Wet season. Blue triangles represent species.

CCA analysis is a powerful ordination technique that allows the illustration of site, species and environmental variable relationships. CCA is an ordination technique that assume a direct unimodal relationship between environmental variables and species data at sample sites. Therefore, the change in diatom community structure across sites can be illustrated with the corresponding environmental variables at sites. The CCA (Figure 30) contains convex hulls for dry and wet seasons, the convex hulls were inserted into the CCA independent of the calculation, therefore we can ordinate the relationship between diatom communities and environmental variables to determine if temporal variability has an underlying influence on the change in community structures. From the CCA, it is evident that the convex hulls for dry and wet season have a large overlap, suggesting the diatom community is correlated more with environmental variables than seasonal change.

From the CCA analysis, pH, alkalinity and sulphate are the main driving factors influencing the species distribution on the first axis with strong correlations (-0.93, -0.79 and 0.65 respectively), that suggests species are separated along the sulphate, pH gradient. The first axis of the CCA represents the alkalinity, sulphate and pH gradient while the second axis represents the chloride and EC gradient. This suggests that the second axis is a secondary driving factor in the species distribution with chloride having a strong correlation with the second axis (0.76) and EC having a slightly weaker relationship (0.61).

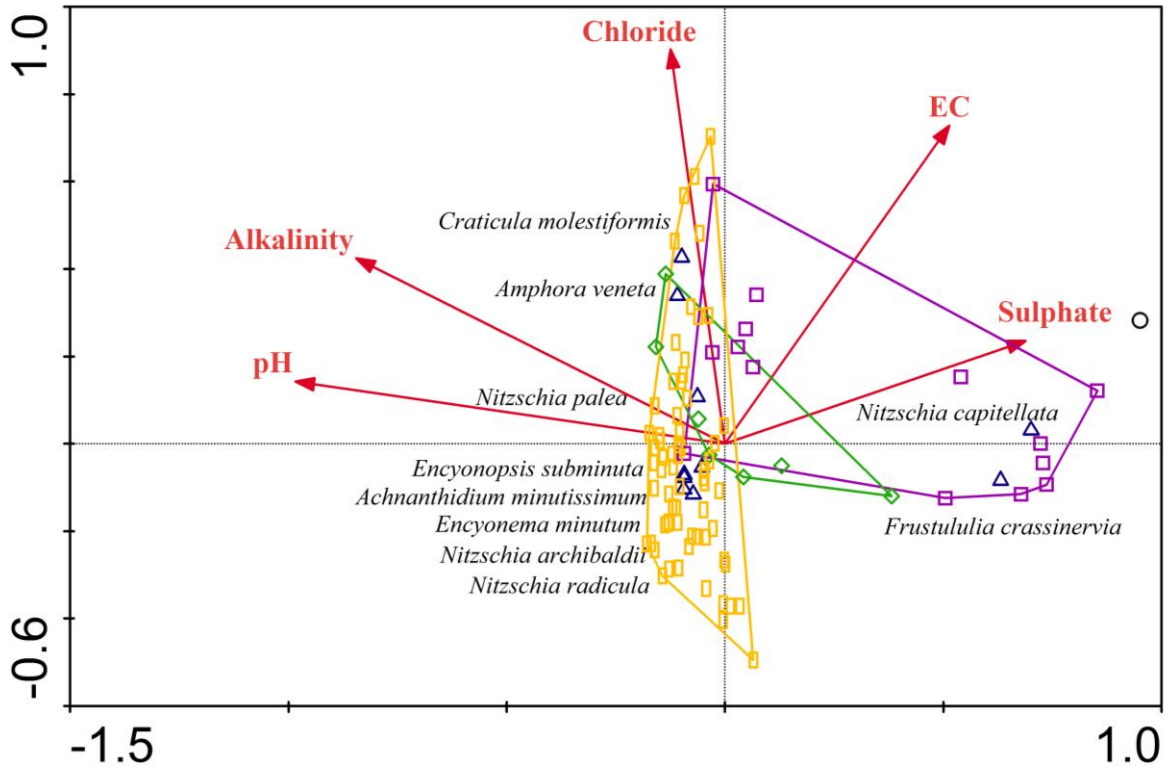


Figure 31: CCA illustrating the relationship between species and environmental variables (red arrows) at sites sampled (geometric shapes). The first two axes explains 12.6% of the variation in species data and accounted for 79% of the species-environment relationship. The convex hulls represent observed disturbance levels of sites (yellow rectangles – low disturbance, green diamonds – moderate disturbance, purple squares – high disturbance, black circle – extreme disturbance). Blue tringles represent species.

Figure 31 illustrates the same species environment relationships at sites sampled, however from the graph it is evident that environmental variables correlate with AMD related disturbance levels (convex hulls). With a higher degree of disturbance, the

concentration of sulphate increase while pH and alkalinity decrease. A weaker relationship exists between chloride and EC with disturbance levels, a weak relationship that suggest as disturbance levels increase the EC and chloride concentrations increase, however as mentioned this relationship is weaker than that of sulphate, pH and alkalinity.

Figure 31 also illustrates taxa contributing to more than 8% of the community structure. *Nitzschia capitellata* and *Frustulia crassinervia* are strongly associated with low pH and increased sulphate and associated with increased EC and decreased alkalinity. These taxa also show a weaker association with chloride concentrations. The associations of these taxa indicate they prefer acidic environments with high sulphate concentrations. Taxa showing a strong association with increased chloride and moderate to high EC are *Amphora veneta* and *Craticula molestiformis*. This suggests these taxa prefer high EC environments mainly determined by chloride concentrations. They are also associated with neutral pH and slightly higher levels of alkalinity.

Other taxa present (*Nitzschia palea*, *Encyonopsis subminuta*, *Achnanthidium minutissimum*, *Encyonema minutum*, *Nitzschia archibaldii* and *Nitzschia radicula*) are abundant in sites with neutral pH and alkalinity, moderate sulphate and moderate conductivity and lower chloride concentrations. As the Sulphate concentrations increase above 800mg/L and the pH decreases below 3, these taxa disappear from the community structure, *N. capitellata* and *F. crassinervia* become more abundant and dominate the community structure in disturbed conditions. As the Chloride concentration and EC increases, the former taxa also disappear from the community and are replaced by *C. molestiformis* and *A. veneta* which become highly abundant.

From the two CCA graphs (Figs 30 & 31) it can also be observed that the species distribution along an environmental gradient, correlates better with observed disturbance levels than seasonal variability. This suggests that seasonal variability has less of an effect in determining diatom community composition than environmental parameter concentrations.

4.3.2) Optimal environmental conditions

The optimum environmental condition for a diatom taxon was only calculated if they occurred in more than ten sites across all samples. When this threshold of occurrence is

set to 10 sites, the influence of outliers on the results is less, the calculation of the optimum condition is more robust, and the influence of rare species is excluded which ensures a more meaningful species-environment relationship since rare species include many zeros during calculation and can distort results and influence the ordination of results.

It is important to note that the relative abundance of a species in the original dataset was used to calculate the optimum environmental condition, the relative abundance was not rescaled with the reduced set of seventy-eight species and still reflects the real relative abundance of a species in the diatom communities sampled.

Using weighted averaging, optimum environmental conditions for taxa were calculated, together with their respective AMD tolerance value, osmotic tolerance value and life-form value (Appendix B). Five key taxa with a strong indication to AMD disturbance in wetlands were identified (Table B.1). These taxa are: *N. capitellata*, *F. crassinervia, Encyonopsis microcephala*, *Nitzschia pura*, *Craticula buderi* and *Navicula veneta*. These taxa have optimum environmental conditions for EC above 1000 µS/cm, sulphate above 350 mg/L, chloride below 150 mg/L an alkalinity below 85 mg/L. Additionally, the pH value for *N. capitellata* and *F. crassinervia* as calculated through weighted averaging was below 4, which indicates these species are correlated with low pH, high sulphate concentrations, high electrical conductivity, low alkalinity and low chloride concentrations. However, of these five taxa only *F. crassinervia* and *N. capitellata* represent more than 8% of the community structure, therefore these species serve as indicators for AMD disturbance in wetlands, where the other three species may serve as potential indicators but do not dominate in AMD disturbed conditions.

Taxa serving as strong indicators for undisturbed wetlands are *Navicula arvensis* var. *maior*, *Amphora veneta*, *Craticula molestiformis*, *Nitzschia amphibia* and *Mayamaea atomus* These taxa all have optimum environmental conditions for pH above 8.5, alkalinity above 130mg/L and chloride above 70mg/L. However, only *A. veneta* and *C. molestiformis* represent more than 8% of the community structure.

Therefore, *N. capitellata*, *F. crassinervia, A. veneta* and *C. molestiformis* can serve as reference taxa for rapid assessment prior to index calculation.

4.3.3) Redundancy Analysis (RDA)

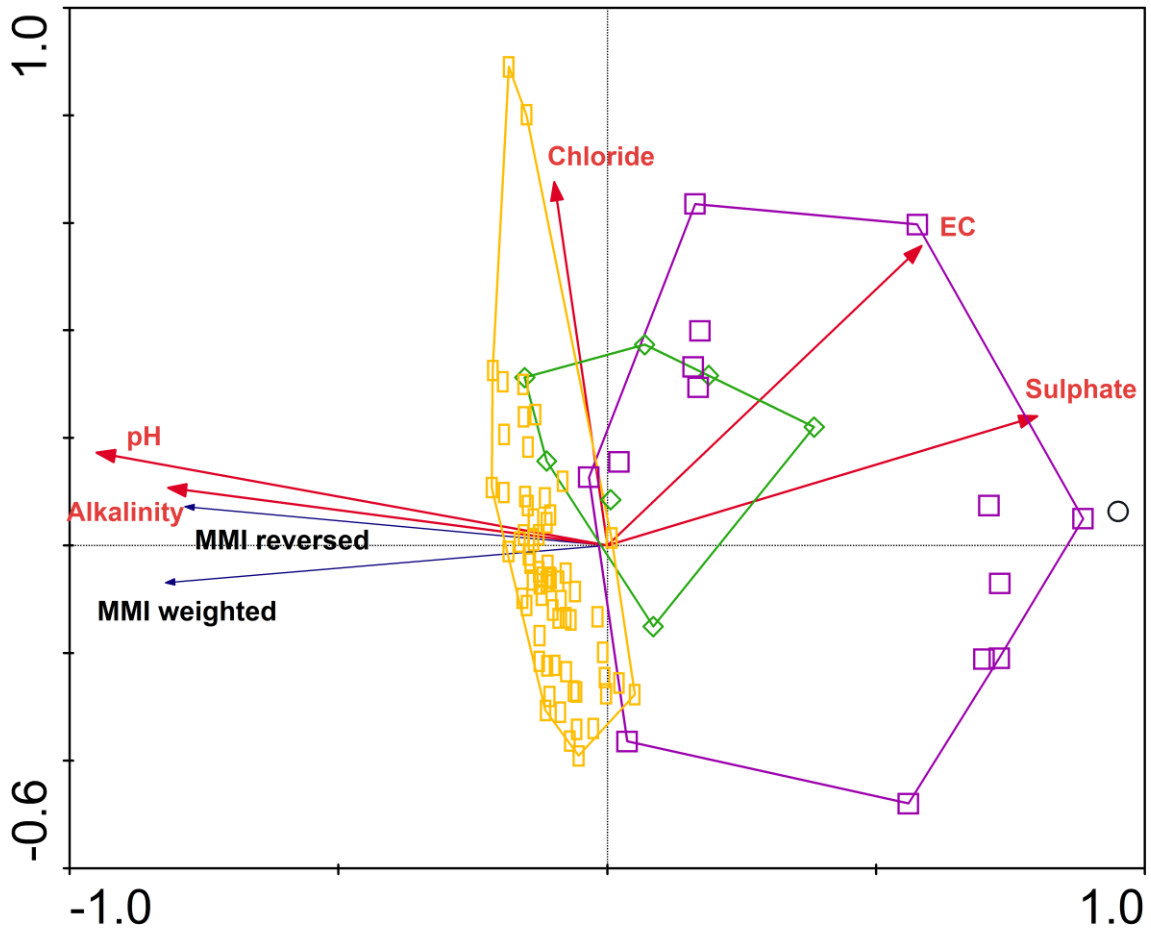


Figure 32: RDA of index scores (blue arrows), environmental variables (red arrows) and sites (shapes). Sites are grouped based on observed disturbances (yellow rectangles – low disturbance, green diamonds – moderate disturbance, purple squares – high disturbance, black circle – extreme disturbance).

An RDA ordination (Figure 32) was chosen to illustrate the relationship between index scores, environmental variables and sites. The five environmental variables used to create the index were used in the RDA to determine how well the calculated scores correlated to the environmental variables associated with AMD impact. An RDA was chosen based on the linear scale of index scores and environmental variables, before the ordination was done environmental data was log transformed.

The RDA shows an increase in the index scores (MMI reversed and MMI weighted) with increasing pH and alkalinity, and deceasing sulphate and EC (Table A.2). The MMI reversed index follows the above methodology without the weighting of environmental variables, whilst the weighted MMI follows the full methodology with weighted

environmental variables; the purpose of this approach is to demonstrate how much influence weighting environmental parameters have on the final index score and correlation. Chloride has a weak positive relationship with the calculated index scores. The index scores also have a negative relationship with disturbance groupings, as the index scores decrease, the disturbance level increases. The MMI reversed has a stronger positive relationship with pH and alkalinity, while the MMI weighted has a stronger negative relationship with sulphate. The RDA indicates that pH and sulphate concentrations are correlated strongly with index scores, as pH increases so does the index score and as sulphate increases, the index scores decrease accordingly.

According to the RDA results, the first axes (x-axes) explain 99.2% of the species-environment relationship and explains 64.6% of the variance in the species data. The first axis explains most of the species environment relationship and is strongly correlated with pH (-0.78), alkalinity (-0.67) and sulphate (0.65). Therefore, the analysis would suggest that the species distribution in sites is highly correlated with pH, alkalinity and sulphate.

4.3.4) Bootstrapping

The reliability and efficiency of indices are validated using independent data or resampling methods (bootstrapping). According to the RDA, the MMI reversed index shows the strongest correlations for the five environmental parameters used to calculate the optimal environmental conditions for species. Additionally, the Pearson correlations of the MMI reversed index with water quality is strong for pH (0.75), sulphate (-0.65) and alkalinity (0.64), weaker for EC (-0.5), and very weak for chloride (0.03). However, these correlations must be validated through resampling.

The calculated optimal environmental conditions for each environmental variable were bootstrapped with error values of 2% of the entire range of measured values and done accordingly: sulphate – 70 mg/L; chloride – 6.5 mg/L; alkalinity – 6.5 mg/L, EC – 15 mS/cm and pH – 0.15. Values below zero were adjusted to 0.1. A total of 1,000 bootstrapping iterations were performed, after which Pearson correlations were calculated for the newly derived indices. The average correlation for sulphate was -0.61 with a 95% confidence interval of [-0.64, -0.57], chloride was 0.08 with a 95% confidence interval of [0.00, 0.17], alkalinity was 0.62 with a 95% confidence interval of [0.57, 0.65],

electrical conductivity (EC) was -0.46 with a 95% confidence interval of [-0.50, -0.39], and pH was 0.72 with a 95% confidence interval of [0.66, 0.75]. The correlations of the index scores with measured water quality variables remained strong and consistent, especially for pH, alkalinity and sulphate, indicating a high level of reliability in the index calculation.

4.5) Discussion

Over the past decade, diatoms have been used to determine the impacts of AMD on wetland systems, however, many of the diatom indices used in South Africa were adapted from European indices designed for lotic, or riverine, systems. The application of riverine diatom indices on wetlands systems is problematic. The main factor causing inconsistent results in diatom index scores is the nature of wetland systems compared to riverine systems. Wetlands are neither lentic, nor lotic and are highly dynamic environments at the intersection of terrestrial and aquatic environments (NWA, 1998). Wetlands experience a varying degree of seasonal rainfall and may be permanent, seasonal or ephemeral. Wetlands are also often characterized by high concentrations of nutrients and organic material with a high rate of sedimentation. Wetlands serve as environmental sinks, trapping and filtering out excess nutrients, organic material, pollutants and toxicants from water before entering riverine systems (Fisher & Acreman, 2004). Therefore, applying riverine diatom indices to a wetland system is inadvisable since these aquatic systems are not comparable in their ecosystem function and spatial variation.

The development of a diatom index to detect AMD disturbance in wetlands must be designed with a particular focus on the nature of wetland systems and how those systems react to AMD disturbance. The diatom indices used in South Africa, adapted from European riverine indices, were developed with a specific focus on detecting eutrophication and organic pollution in lotic systems from agricultural and urban runoff, and wastewater discharge. These indices were developed with specific attention on nutrient enrichment and organic pollution. Wetlands have a naturally high concentrations of nutrients and organic material and therefore the use of riverine indices will result in many wetland systems being classified as highly impacted systems although it may not be the case.

Wetland systems are adversely affected by AMD disturbance. Wetlands have a natural assimilative capacity for AMD with minimal impact to the environment, however, excessive contamination through AMD can have significant effects on their ecological function. When the natural assimilative capacity is exceeded, the acidification of water occurs, increased concentrations of heavy metals and toxic compounds increase as pH decreases and the ecosystem resilience and function decreases (Tatu *et al*., 2008; Dean *et al*., 2013). It is worth noting that AMD disturbance in wetlands caused by gold/uranium mines is more of a concern than AMD disturbance from coal mines. AMD effluents from gold/uranium mines are highly acidic with high concentrations of heavy metals that can adversely affect wetlands for decades (Laker, 2023). On the other hand, AMD from coal mines is characterized by high sulphide (S^{2-}) and sulphate (SO_4^2) concentrations with fewer traces of heavy metals than AMD from gold/uranium mines. AMD from coal mines is also manageable through the application of lime, to neutralize the acidity of AMD effluents for example. (Mogashane *et al*., 2025).

AMD from coal mines contains high concentrations of sulphide and sulphate compounds that increase the acidity of surrounding water and increase the solubility of heavy metal compounds. The oxidation of pyrite forms sulfuric acid that lowers the pH within wetland systems. When excessive amounts of sulphide and sulphate compounds enter wetlands, the buffering capacity is reduced, and the system becomes increasingly acidic (Tatu *et al*., 2008; Pat-Espadas *et al*., 2018). Chloride concentrations can also increase in wetlands when impacted by AMD, however, high chloride concentrations are not a direct effect from AMD but rather chloride compounds leaching into the environment following mining activities and from urban runoff. High chloride concentrations in wetlands are not an indicator of AMD, rather a metric for natural weathering (Annandale *et* al., 2002; Gebrekristos & Trusler, 2018). The ratio of sulphate to chloride can therefore provide valuable insights into the source of pollution. A high sulphate to low chloride ratio would indicate AMD disturbance from mining activities while a low sulphate to high chloride ratio would indicate natural ageing of wetlands, especially those with endorheic flow (Humphries *et al*, 2017). Alkalinity is also an important metric to consider when determining the impact of AMD on wetland systems. Alkalinity is a measure of the buffering capacity of water to prevent acidification. When the alkalinity is high in wetlands,

the effects of AMD are less pronounced due to the buffering capacity of the wetland system (Mayes *et al*., 2009; Lusunzi *et al*., 2023). Low alkalinity would therefore place a wetland system at increased risk of acidification should there be an AMD disturbance. The increase in acidity in wetlands systems often correlates with a high dissolution of heavy metals, mineral salts and pyrite (Woulds & Ngwenya, 2004; Grantcharova & Fernández-Caliani, 2021). This increases the ionic load in the water and consequently increases the electrical conductivity (EC). EC is a valuable metric to consider when creating a diatom index for AMD disturbance since electrical conductivity is one of the most important and highly correlated environmental factors influencing diatom community structure (Dalu & Froneman, 2016; Fernández-Moreno *et al*., 2024).

These five environmental parameters (sulphate concentration, pH, alkalinity, chloride concentration and EC) characterize AMD disturbance very well. With increased AMD disturbance, sulphate concentrations increase, pH decreases, alkalinity decreases, chloride concentrations decrease and EC increases. These five parameters were therefore used as metrics in the calculation of the newly proposed multimetric index.

The calculation of the index is a combination of optimal environmental conditions, AMD and osmotic tolerances, and life form values determined for 78 species. The abundance of species at sites serves as additional weights in the calculation of the index score. However, only 10 species represent more than 8% of the community structure and therefore contribute most to the calculated index scores. These index scores correlate well with water quality (pH = -0.79, alkalinity = -0.68, sulphate = 0.62), suggesting these species respond to water quality changes and can be used as indicator species of AMD disturbance in wetlands.

According to the results, unimpacted wetlands are dominated by *Achnanthidium minutissimum*, *Encyonopsis subminuta*, *Encyonema minutum*, *Nitzschia archibaldii*, *Nitzschia radicula* and *Nitzschia palea*. These species share a diversity of life-forms (motile, attached and tube-forming). As water quality shifts to higher sulphate concentrations, low pH, low alkalinity and elevated EC, the community structure shifts to one dominated by *Nitzschia capitellata* and *Frustulia crassinervia*. This suggests that these species are indicators of AMD impacted wetlands as their dominance in the

community structure correlates with observed disturbance levels. These species are motile and are known to occur under highly acidic environments (Passy, 2007; De Nicola, 2000; Kulichová & Fialová, 2016).

As water quality shifts to higher chloride concentrations and elevated EC, the community structure is largely dominated by *Amphora veneta* and *Craticula molestiformis* which are small, motile species. This suggests that these species are indicative of unimpacted wetlands with increase chloride concentrations and are known to occur in electrolyte rich waters (Taylor *et al*., 2007c). These species are also small, motile species. The taxa that dominate the community structure are therefore a mix of different life forms, however, motile species are present in all disturbance levels, suggesting that life form is not a diagnostic trait that can be used to assess AMD impacted wetlands, rather, the occurrence of these species are likely linked to physiological adaptations. It is also worth noting that diatom teratology's were not considered in this study but could be an important morphological feature to consider in future studies as they are linked to metal stress and can be indicators of AMD disturbance.

Attempts to use diatoms as a biomonitoring technique for AMD disturbance in wetlands were initially not particularly successful, largely due to the use of riverine indices in wetland systems without any modification to the calculation of indices. Recent studies by Riato (2017) have attempted to create a multimetric diatom index for use in wetlands and have been largely successful, however, the main concern with the method developed by Riato (2017) is its complexity and time-consuming nature in terms of calculating the index scores. The method provides accurate results; however, the complex nature of the index requires users to have in depth knowledge of the structure of diatom communities, statistical analysis and mathematical formulas. The use of this method by consultants and others involved in the routine monitoring of water resources is therefore not feasible as a reliable, quick way to determine AMD impacts in wetland systems. Keeping in mind the complexity of multimetric indices and the importance for accurate, easily obtainable results, the present study proposes a refined multimetric index for AMD detection in wetlands using diatoms. The index calculation is housed in an executable computer programme where index scores are automatically calculated from an input sheet. Therefore, the index could be valuable to the Department of Water and Sanitation (DWS),

the South African National Biodiversity Institute (SANBI) and provincial management authorities that assess wetlands impacted by mining activities, by providing an easy-to-use platform that can increase data generation and throughput, and thereby supporting wetland health and rehabilitation, and policy implementation.

4.6) Summary

The unique nature of wetlands systems as compared to riverine systems requires special attention when creating water quality indices to detect AMD. Using riverine indices is inadvisable since they are specifically designed to detect eutrophication and organic contamination and would classify healthy wetlands as impacted. Creating a multimetric diatom index to detect AMD based on sulphate, pH, alkalinity, EC, and chloride is more desirable since these five criteria characterise AMD in wetlands. By using a multimetric approach, the optimum environmental conditions for diatom taxa for five key parameters were determined. Individual indices for each parameter at each site were then calculated, followed by a weighted average of these indices. This method effectively distinguished between AMD-impacted and non-impacted wetlands. RDA illustrated that the species distribution in sites is highly correlated with pH (-0.78), alkalinity (-0.67) and sulphate (0.65). Resampling using bootstrapping also confirmed these correlations and serves as a validation for the reliability of the index. Additionally, four key taxa were identified as early indicators for AMD disturbance – *Nitzschia capitellata* and *Frustulia crassinervia*, *Amphora veneta* and *Craticula molestiformis*. Therefore, this multimetric approach was successful in detecting AMD disturbance in wetlands surrounding coal mines and can be used by government and provincial management authorities to assess wetlands impacted by mining activities.

CHAPTER 5: DIGITAL TOOLS TO FACILITATE DIATOM INDEX CALCULATION

5.1) Introduction

Creating indices and generating diatom index results had been done manually before the introduction of the OMNIDIA software program. Before the introduction of computer software, calculating index scores was tedious and time consuming. The throughput of data generated is increased when introducing index calculation software that calculates index scores and well as archiving scores with recorded species counts. OMNIDIA remains the most widespread version of the software and has been in use since its release in 1993 (Lecointe *et al*., 1993). OMNIDIA offers a robust analysis of diatom data using multiple indices, including the IPS, GDI, TDI, BDI and many more. The output of the software is sufficient; however, the software can be difficult to navigate when analysing data, and the input of data is complex and requires time consuming preparation from the user. The licensing of the OMNIDIA software can be restrictive since only those able to pay for the licence can use the software and this factor can potentially diminish data output. The OMNIDIA software is also run locally on computers, and when these computers are lost or broken, the data is also lost. Additionally, no online repositories for diatom index scores and site data exist for South Africa. Therefore, the current use of diatom index software is limited by OMNIDIA and the limitations of difficulties of use of the software, the dissemination of data generated is diminished through locally stored data repositories and licensed operation of the software. Furthermore, the OMNIDIA software does not house an index calculation for wetland systems, only riverine indices. However, it also houses Japanese and South American indices.

Herein new digital tools are proposed that generate index scores and archive sites scores and species abundances without the need for a licence. A Diatom Indexer is proposed that houses the calculation of the most reliable riverine indices proposed in present study as well as the SADI index as modified by Harding & Taylor (2011) from the IPS. Additionally, a simple executable software program is provided to calculate index scores for wetlands impacted by AMD disturbance as proposed in section (4.2) .

5.2) Methods

Section 3.2.1 explains the different methods used to calculate species optima and tolerance values, combining those values into index calculations (Sections 3.2.2 and 4.2) and rescaling the index values to an interpretable format (Sections 3.2.3). This process cannot be repeated every time a new index is calculated. Therefore, the calculated optima and tolerance values are recorded into a dataset where they are housed and linked to a specific species code. Additionally, the species optima and tolerance values used in the IPS are also housed in a similar way and linked to a species code. These datasets can be used in an executable software program as part of the backend to calculate index scores for users. Users provide input data in the form of species matrix with species relative abundances at sites. The species code is linked to the optima and tolerances housed in the backend and an index value is generated for the user together with a graph that visualizes the index scores across all sites given and for each index calculated. Additionally, species ecology is provided for species in the input dataset for better interpretation end elucidation of index scores.

The Diatom Indexer includes a front-end interface where users can create new projects with site names, river names, site locations and date. To simplify the input of data, the same format used by OMNIDIA software is also used by the diatom indexer, therefore, users need not alter their historical data already structured to fit the OMNDIA input format and can simply input this data into the program. Users can also access older projects through an ‘open project’ button (Figure 33). Additionally, a help section is included that explains the indices in the software and explains how to navigate the program.

The executable software that calculates AMD disturbance for wetlands does not include a front-end interface. Instead, a downloadable folder containing the program is needed. Users are required to copy their input file, a simple species plot matrix, into the relevant folder and run the program. An excel spreadsheet is generated that contains the calculated index scores.

5.3) Results

5.3.1) Diatom Indexer

The Diatom Indexer was designed for the calculation of riverine indices proposed in this dissertation as well as the SADI as modified by Harding & Taylor (2011) from the IPS.

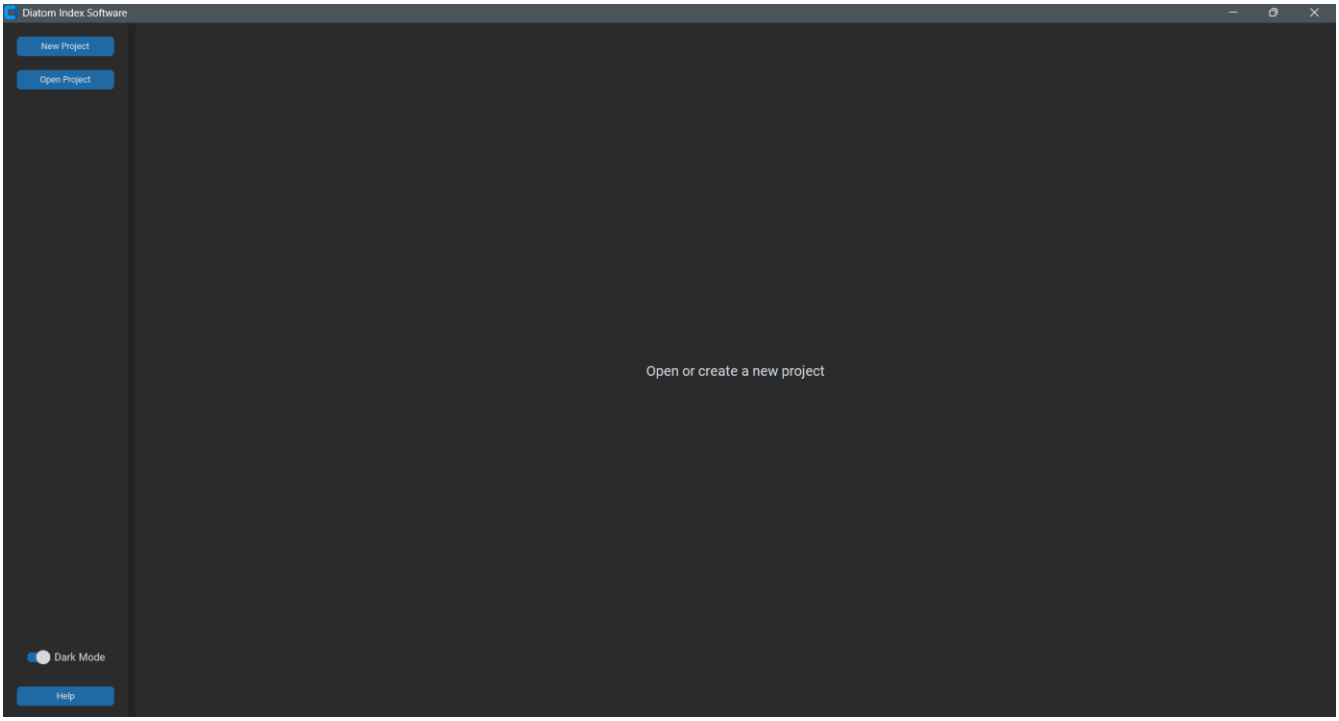


Figure 33: The landing page of the Diatom Indexer software.

In Figure 33 the initial landing page of the Diatom Indexer software is displayed. The user interface is simple and only temporarily includes the necessary options to create new projects, open previous projects (for the time being all projects are stored locally on the user's computer and can be accessed through by the 'Open Project' button (Figure 35), and a help section that explain how to use the program and explains the diatom indices displayed in the results section.

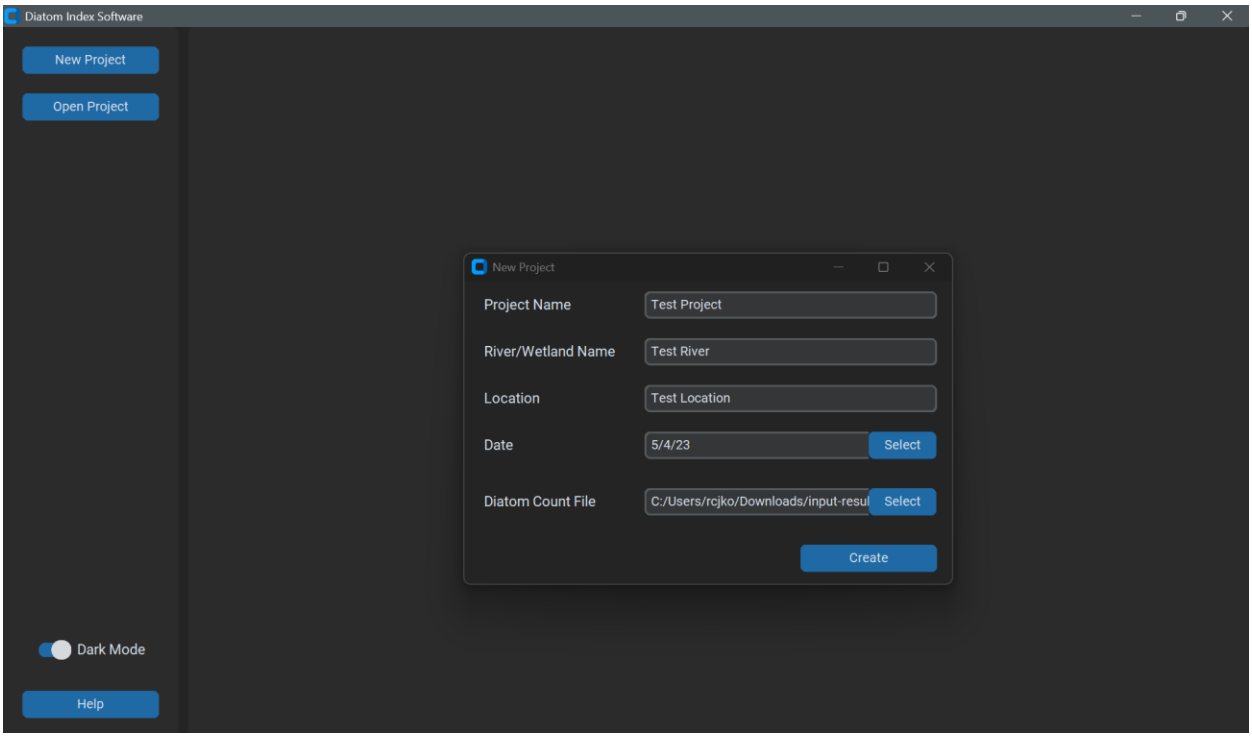


Figure 34: Input requirements for a new project.

Figure 34 displays the information needed to create a new project. Users are required to input a project name, a river or wetland name, a location or locations for the site data (text) as well as a date for sampling or score generation. Users are required to select a diatom count file. The Diatom count file is a .csv file and has the same format as required by OMNIDIA v 5.3. Once the user creates the project, the diatom index scores are automatically calculated.

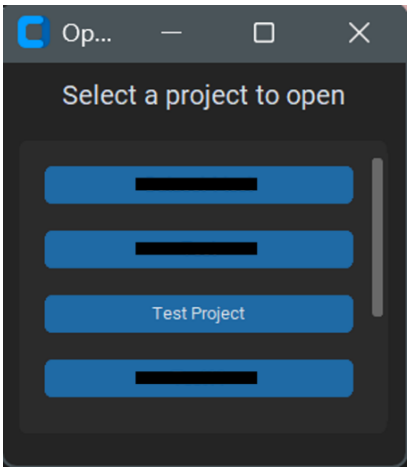
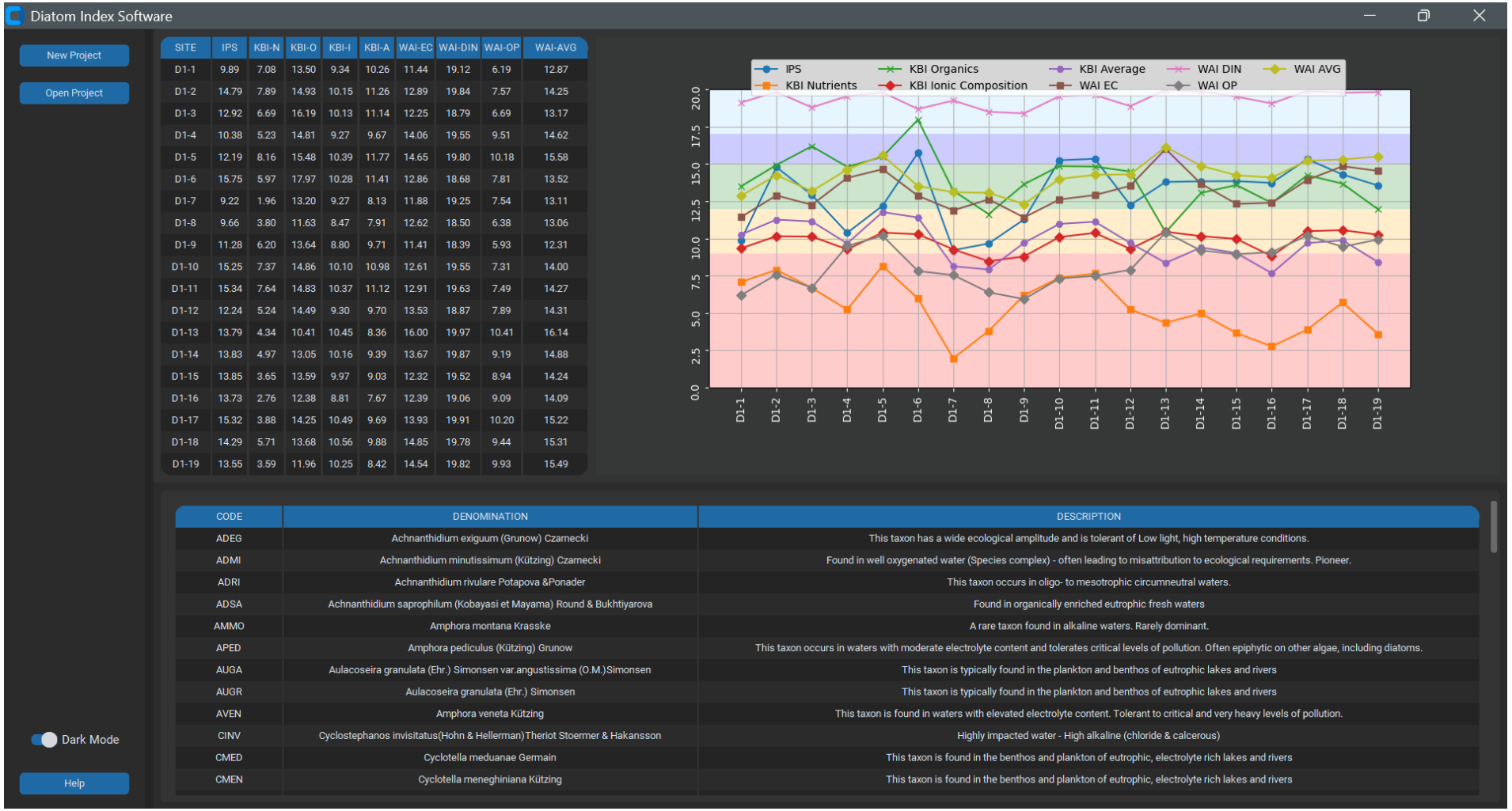


Figure 35: Example of the 'Open Project' window.



SITE	IPS	KBI-N	KBI-O	KBI-I	KBI-A	WAI-EC	WAI-DIN	WAI-OP	WAI-AVG
D1-1	9.89	7.08	13.50	9.34	10.26	11.44	19.12	6.19	12.87
D1-2	14.79	7.89	14.93	10.15	11.26	12.89	19.84	7.57	14.25
D1-3	12.92	6.69	16.19	10.13	11.14	12.25	18.79	6.69	13.17
D1-4	10.38	5.23	14.81	9.27	9.67	14.06	19.55	9.51	14.62
D1-5	12.19	8.16	15.48	10.39	11.77	14.65	19.80	10.18	15.58
D1-6	15.75	5.97	17.97	10.28	11.41	12.86	18.68	7.81	13.52
D1-7	9.22	1.96	13.20	9.27	8.13	11.88	19.25	7.54	13.11
D1-8	9.66	3.80	11.63	8.47	7.91	12.62	18.50	6.38	13.06
D1-9	11.28	6.20	13.64	8.80	9.71	11.41	18.39	5.93	12.31
D1-10	15.25	7.37	14.86	10.10	10.98	12.61	19.55	7.31	14.00
D1-11	15.34	7.64	14.83	10.37	11.12	12.91	19.63	7.49	14.27
D1-12	12.24	5.24	14.49	9.30	9.70	13.53	18.87	7.89	14.31
D1-13	13.79	4.34	10.41	10.45	8.36	16.00	19.97	10.41	16.14
D1-14	13.83	4.97	13.05	10.16	9.39	13.67	19.87	9.19	14.88
D1-15	13.85	3.65	13.59	9.97	9.03	12.32	19.52	8.94	14.24
D1-16	13.73	2.76	12.38	8.81	7.67	12.39	19.06	9.09	14.09
D1-17	15.32	3.88	14.25	10.49	9.69	13.93	19.91	10.20	15.22
D1-18	14.29	5.71	13.68	10.56	9.88	14.85	19.78	9.44	15.31
D1-19	13.55	3.59	11.96	10.25	8.42	14.54	19.82	9.93	15.49

CODE	DENOMINATION	DESCRIPTION
ADEG	Achnanthidium exiguum (Grunow) Czarnecki	This taxon has a wide ecological amplitude and is tolerant of Low light, high temperature conditions.
ADMI	Achnanthidium minutissimum (Kützing) Czarnecki	Found in well oxygenated water (Species complex) - often leading to misattribution to ecological requirements. Pioneer.
ADRI	Achnanthidium rivulare Potapova &Ponader	This taxon occurs in oligo- to mesotrophic circumneutral waters.
ADSA	Achnanthidium saprophilum (Kobayasi et Mayama) Round & Bukhtiyarova	Found in organically enriched eutrophic fresh waters
AMMO	Amphora montana Krasske	A rare taxon found in alkaline waters. Rarely dominant.
APED	Amphora pediculus (Kützing) Grunow	This taxon occurs in waters with moderate electrolyte content and tolerates critical levels of pollution. Often epiphytic on other algae, including diatoms.
AUGA	Aulacoseira granulata (Ehr.) Simonsen var.angustissima (O.M.)Simonsen	This taxon is typically found in the plankton and benthos of eutrophic lakes and rivers
AUGR	Aulacoseira granulata (Ehr.) Simonsen	This taxon is typically found in the plankton and benthos of eutrophic lakes and rivers
AVEN	Amphora veneta Kützing	This taxon is found in waters with elevated electrolyte content. Tolerant to critical and very heavy levels of pollution.
CINV	Cyclostephanos invisitatus(Hohn & Hellerman)Theriot Stoermer & Hakansson	Highly impacted water - High alkaline (chloride & calcerous)
CMED	Cyclotella meduanae Germain	This taxon is found in the benthos and plankton of eutrophic, electrolyte rich lakes and rivers
CMEN	Cyclotella meneghiniana Kützing	This taxon is found in the benthos and plankton of eutrophic, electrolyte rich lakes and rivers

Figure 36: The main User Interface (UI) of the diatom indexer program show Index calculation results (top left), a graph illustrating trends in the index scores (top right) and a list with species ecological descriptions (bottom).

Figure 36 displays the index results as generated for the input project. For illustration purposes, the first 19 sites in the first dataset are displayed in the software (sites D1 – D19) (Table A.4). A table with index scores is included (left) where users can view index scores, together with a graph that illustrates site index scores calculated for all indices included in the software (right), as well as some ecological data for species used to generate the index score (bottom). The graph displayed is coloured, based on the ecological classes in Table 7 and allows the user to identify and follow a trend in the index scores across sites. Additionally, the individual indices for environmental parameters (EC, DIN and PO_4^{3-}) can be tracked and used to identify a particular environmental impact. For example, the index scores for site D1-9 had slight variation among the different weighted average indices. The WAI-AVG suggests a mesotrophic state, while the WAI-EC score reflects a eury-mesohaline condition, and the WAI-DIN score indicates an oligotrophic state. The overall WAI-AVG score is reduced by the WAI-OP value, which shows eutrophic conditions. Therefore, although site D1-9 is classified as mesotrophic (approaching the meso-eutrophic boundary), this is primarily due to elevated orthophosphate levels.

Together with the index scores (the table and graph), users can consult the ecology table given below the graph. This table contains all the species codes as listed in the user's input file. Therefore, when analysing site scores, users can consult this table to gain addition information on the interpretation of their results.

5.3.2) AMD index executable software

The AMD index calculation software is presented as a standalone executable program. Figure 37 displays the contents of the downloadable folder. An archive file containing the executable program is given with the files folder. Once extracted the AMD-indexer.exe programme is available in the folder. The 'files' folder contains a standard.xlsx file (Figure 38) which houses all the optima and tolerance values generated for species as listed in Table (B.1) and contains the life forms as well as the AMD and osmotic tolerances for species.

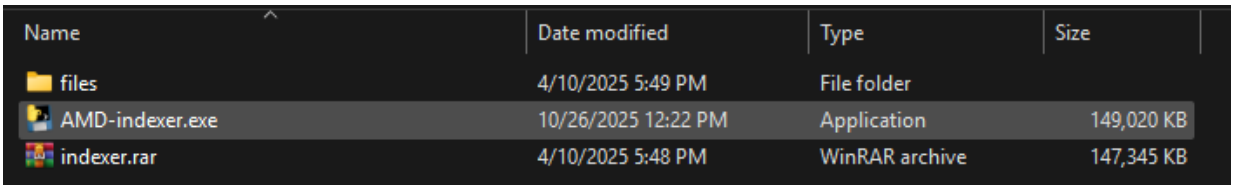


Figure 37: Content of the AMD indexer folder.

The input file is placed within the 'files' folder and should be renamed to 'input.xlsx' and is a standard species matrix with sites in the top row and species relative abundances in coloumns.

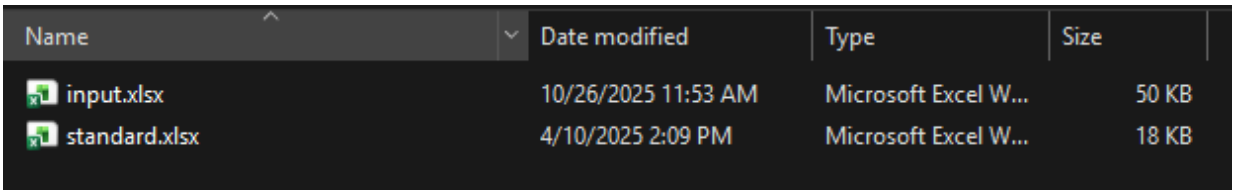


Figure 38: Content of the files folder.

Once the input file is placed into the files folder, the AMD-indexer.exe is run and a index_summary.xlsx file is generated that contains the index scores calculated for sites based on section (4.2) (Figure (39).

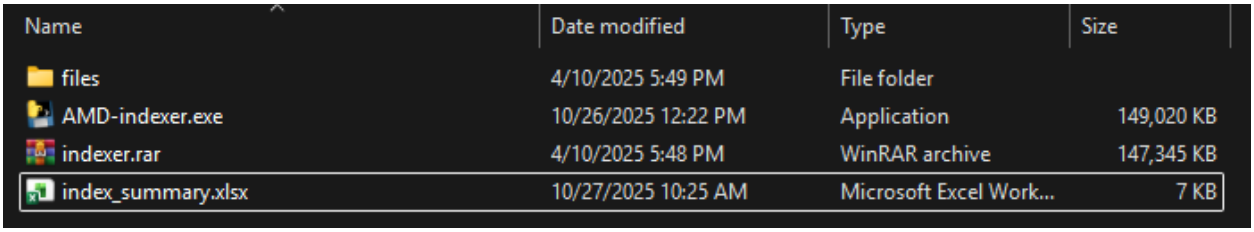


Figure 39: The updated content of the AMD indexer folder after index calculation.

5.4) Discussion

Digital innovation has become the new standard in the 21st century, and with the introduction of AI in recent years, the amount of information housed and generated online is increasing. Ecology has been the recipient of many advancements due to digitial innovation, (e.g. iNaturalist). The use of AI, in particular machine learning, has aided ecologists in the statistical analysis of results to better understand environemntal relationships and to make increasingly accurate predictions. Digital innovation in the field of biomonitoring has seen an increased interest, however, is still in the early stages of incorporating digital tools. In present study, new digital tools are proposed for the

calculation of diatom indices for riverine envirnmenat and wetland systems. The Diatom Indexer was effecient in calculating index scores using standard lists as a reference, these standard lists house the optima and tolerance ranges as calculated using WA and inferred from expert knowledge, in comparison with the IPS. The calculation of the IPS scores corresponds with the output of the OMNDIA software and can therefore serve as replacement for the calculation of the IPS. The main purpose of the Diatom Indexer is to simplify the calculation of diatom indices and to present the results in an interpretable format. The user interface in the results page of the Diatom Indexer is sufficient for illustrating the index scores calculated, illustrating trends in diatom index scores across sites and comparing different indices in an interpretable format, as well as providing the ecology data for many species to further aid in the interpreation of index scores. The backend of the software, used to calculate indices, works well, however, the user interface can be updated to streamline the process more and to provide additional tools to aid in the interpretation of index scores. Currently the user interface demonstrates the possibilities of digital innovation with diatom software, especially for South Africa. It offers an alternative to the OMNDIA software for South African diatom researchers as a simplified, easy-to-use and free software.

The Diatom Indexer still has many changes needed to make it an effecient tool for routine monitoring in South Africa. Currently, the metadata for a new project only inlcudes a project name, a river or wetland name, a location (text) and a date of sampling or processing. Other information such as site images and catchment locality can be added to the broader project classification. For indiviual sites in the dataset, site data can be added such as site name, coordinates, measured water quality, elevation, etc. This will strenghten the usability of the project data for repititon and routine monitoring.

Furthemore, the software is currently run locally on personal computors, and data is not automatically stored in an online repository. Enabling storage online can improve the safekeeping of data generated and allow users to access shared data for holistic projects concerning water quality monitoring in South Africa. This is however difficult to achieve without the neccasary funding and maintainenance budget to create and manage an online repository with a purchased domain. However, when such funding is availabile and an online repository is created, the possibilties of improving the Diatom Indexer is further

increased with the inclusion of a regional map that indicates sample sites and diatom index scores for those sites. Additionally, if each sites is accesed and index scores are displayed across sampling years, the information can be used to identify trends in water quality degredtion or improvement and can help to manage water qaulity goals as set out in the NWRS.

The indexing software for AMD index calculation also provides an easy to use and free software for the calculation of AMD indices. However, this software is much simpler than the diatom indexer for riverine indices. The calculation of the AMD index differes from the calculation of the riverine indices where all the indices are calculated using the formula proposed by Zelinka and Marvan (1961), the AMD index uses a different novel calculation formula and inlcuding this into the Diatom Indexer is complex and requires extenisve coding. Furthemore, the Diatom Indexer software has no indication for which system (riverine or wetland) the index must be calcuated using the input file. Therefore to incorporate this operation, the UI must be changed to facilitate the different input files. Before creating a project, the user must select the aquatic system in which the data was generated and from there seperate calculations can be done and displayed separately. To simply include the AMD index scores in the UI as it is represented currently is not accurate since a riverine dataset cannot have an AMD score and similarly a wetland dataset cannot have a riverine index score. The inclusion of the AMD index in the Diatom Indexer will further improve the capability of the software to efficiently be used as a biominitoring tool for multiple aquatic systems. Including AMD results in the map as described above will further improve the information given to managers to improve descision making around water quality goals.

5.5) Summary

The Diatom Indexer software for both riverine indices and the AMD index calculation tool for wetland systems are efficient in calculating and illustrating index scores in an interpretable way, however, the digital innovation of these tools can be improved through changes to the UI, communication with an online data respository, a site map and the inclusion of riverine and AMD indices in one software package. This will improve the efficiency of diatoms as biomonitoring tools in South Africa by providing a free to use tool

to aid in the storage, throughput and analysis of data to improve managment descisions regarding water quality changes in South Africa.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

In South Africa, where water quality issues are widespread, monitoring water quality using biological organisms is widely implemented. However, some water quality issues remain unreported and inaccessible from scientific research. This is where the newly designed digital tools proposed in this study can contribute to mapping water quality issues and providing researchers and citizens with robust tools for water quality assessments. The purpose of creating the digital tools for diatom indices and miniSASS surveys is to improve data storage, generation and throughput for analysis and to improve the effective dissemination of information to management level, thereby supporting better decision-making regarding water quality goals.

The development of the miniSASS application was successful in providing users with an easy-to-use, streamlined tool for miniSASS assessments while disseminating information on water quality monitoring and biological surveys. However, it is recommended that the machine learning model be retrained using more images, at family level instead of group level. Using family level classification and then combining all families belonging to the same higher taxonomic group as used in miniSASS will improve the classification of macroinvertebrates by averting the high morphological diversity within groups. This adjustment would help alleviate the bottleneck of manual verification of site scores as well as introducing the use of machine learning for SASS5.

Modifications to the miniSASS site page in the application are recommended to include site images, sampling dates, and photos for each site assessment. Furthermore, the map page in the mobile application significantly slows down overall performance. Fast response times are essential, as delays can discourage users from continuing to use the app. The map page also lacks sufficient user interaction and is currently severely underutilised. The digital identification must also be revised, to ensure unique identifiers for each group and to refine which identifiers are used.

The diatom indexing software for riverine indices and AMD disturbance offers efficient and accurate calculation of index scores. However, further improvements are necessary to streamline the index calculation process. Firstly, larger datasets must be generated

and used to recalculate species optima and tolerance values and must also be calculated for more species. This study presented Weighted Averaging (WA) and inferred knowledge as ways of calculating species optima and tolerances for riverine diatoms. The success of the optima and tolerance calculations is demonstrated with many values corresponding to the IPS values.

The IPS index, being the most widely used diatom index in South Africa accurately reflects river water quality, and provides a benchmark for comparison; pairing it with the indices developed in this study allows identification of discrepancies and highlights areas where the new indices can be refined and improved. The indices generated using WA and inferred knowledge correlate well with the IPS and importantly with overall water quality. Indices calculated for individual parameters indicate specific water quality degradation while the IPS and the combined indices indicate overall water quality. Furthermore, the optima and tolerances for species not currently in the IPS can be calculated using WA and included in the IPS for better use in the context of South Africa. However, species optima and tolerances can only be included if calculated by using large datasets to ensure reliability. It is therefore recommended that datasets used to calculate species optima and tolerances using WA be expanded to more accurately capture water quality variations and ecological responses of species.

Demonstrating the optima and tolerances for species sampled in wetlands is more difficult since no documented optima exists in South African literature. The reliability of these optima is however bolstered by the index correlations confirmed with bootstrapping. However, it is also recommended that larger datasets be used to calculate species optima for wetland diatoms as well. Furthermore, the inclusion of life-forms and guilds as weights in the index calculation may prove unnecessary, since these adaptations reflect physical habitat and not necessarily responses to water quality. When calculating species optima for wetland diatoms, their physiological responses to specific stressors should be used, since this provides a more accurate reflection of diatom response to wetland health.

Nevertheless, the AMD indexing software was successful in calculating index scores for wetland systems impacted by AMD disturbance. However, the same recommendation applies, larger datasets are needed to calculate optima and tolerance values for species

for AMD disturbance to more accurately reflect AMD-related impacts. The presentation of AMD index results can be improved by incorporating the AMD index calculation in the Diatom Indexer software to improve functionality and the efficiency of digital tools housing diatom indices by providing a single tool for calculating riverine and wetland indices. Additionally, the software should automatically store results in an online repository to ensure data safekeeping and include an online map displaying site scores. These enhancements will allow the diatom indexer software to replace the OMNIDIA software by offering a free, user-friendly platform that provides robust analysis of diatom community assembles as reflected by water quality changes in riverine and wetland systems.

Overall, the novel digital tools developed for diatom indices and miniSASS provide effective platforms for disseminating water quality information and conducting reliable surveys that support the generation of larger, more consistent datasets. By empowering citizens, expanding public participation, and increasing both public and scientific awareness, these tools extend the reach of biomonitoring beyond traditional expert-led programmes. The recommended online repositories will further enhance data storage and accessibility, ensuring that index calculations are efficient, standardised, and less vulnerable to data loss. With more extensive, reliable datasets and broader engagement, the management goals outlined in the NWRS can be more effectively achieved, helping South Africa move steadily toward fulfilling the objectives of SDG6.

REFERENCES

Adams, J., Taljaard, S., van Niekerk, L., & Lemley, D. 2020. Nutrient enrichment as a threat to the ecological resilience and health of South African microtidal estuaries. *African Journal of Aquatic Science*, 45(1–2):23-40.

Adom, R.K. & Simatele, M.D. 2024. Overcoming systemic and institutional challenges in policy implementation in South Africa's water sector. *Sustainable Water Resources Management*, 10(69).

Akcil, A., & Koldas, S. 2006. Acid Mine Drainage (AMD): causes, treatment and case studies. *Journal of cleaner production*, 14(12-13):1139-1145.

Alexander, R.B., Boyer, E.W., Smith, R.A., Schwarz, G.E. & Moore, R.B., 2007. The role of headwater streams in downstream water quality. *Journal of the American Water Resources Association*, 43(1):41-59.

Alizadeh, O. 2022. A review of the El Niño-Southern Oscillation in future. *Earth-Science Reviews*, *235*:104246.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. & Farhan, L. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1):53.

Amoah, I. D., Kumari, S., Reddy, P., Stenström, T. A., & Bux, F. 2020. Impact of informal settlements and wastewater treatment plants on helminth egg contamination of urban rivers and risks associated with exposure. *Environmental Monitoring and Assessment*, 192(11):713.

Angelakis, A. N., Capodaglio, A. G., Passchier, C. W., Valipour, M., Krasilnikoff, J., Tzanakakis, V. A., Sürmelihindi, G., Baba, A., Kumar, R., Haut, B., Roubelakis, M. G., Min, Z., & Dercas, N. 2023. Sustainability of water, sanitation, and hygiene: From prehistoric times to the present times and the future. *Water*, 15(8):1614.

Ansari, F., Pandey, Y. K., Kumar, P., & Pandey, P. 2013. Performance evaluation of effluent treatment plant for automobile industry. *International Journal of Energy and Environment,* 4(6):1079-1086.

Arimoro, F.O., Masese, F.O. & O'Brien, G.C. 2024. Advances in biomonitoring of African aquatic ecosystems, volume II. *Frontiers in Water*, 6:1516711.

Ärje, J., Melvad, C., Jeppesen, M.R., Madsen, S.A., Raitoharju, J., Rasmussen, M.S., Iosifidis, A., Tirronen, V., Gabbouj, M., Meissner, K. & Høye, T.T. 2020. Automatic image-based identification and biomass estimation of invertebrates. *Methods in Ecology and Evolution*, 11(8):922-931.

Armitage, N. P., Winter, K., Spiegel, A., & Kruger, E. 2009. Community-focused greywater management in two informal settlements in South Africa. *Water Science and Technology*, 59(12):2341-2350.

Baker, E., Drury, J.P., Judge, J., Roy, D.B., Smith, G.C. & Stephens, P.A. 2021. The verification of ecological citizen science data: current approaches and future possibilities. *Citizen Science: Theory and Practice*, 6(1).

Balázs, B., Mooney, P., Nováková, E., Bastin, L. & Arsanjani, J.J. 2021. Data quality in citizen science. *The Science of Citizen Science*, 139:978-3.

Balgah, R. A., Ngwa, K. A., Buchenrieder, G. R., & Kimengsi, J. N. 2023. Impacts of floods on agriculture-dependent livelihoods in Sub-Saharan Africa: An assessment from multiple geo-ecological zones. *Land*, 12(2):334.

Barnard, J., de Clercq, W. & van Rensburg, L. 2012. Case studies from Water Research Commission projects along the Lower Vaal, Riet, Berg and Breede Rivers. *Water SA*, 37(5):739-749.

Bate, G.C., Adams, J.B. & an der Molen, J.S. 2002. *Diatoms as indicators of water quality in South African river systems*. WRC Report No. 814/1/02. Pretoria: Water Research Commission.

Bate, G.C., Smailes, P.A. & Adams, J.B. 2004. Benthic diatoms in the rivers and estuaries of South Africa. WRC Report No. TT234/04. Pretoria: Water Research Commission.

Battarbee, R.W., Jones, V.J., Flower, R.J., Cameron, N.G., Bennion, H., Carvalho, L. & Juggins, S. 2001. Diatoms. In: Smol, J.P., Birks, H.J.B. & Last, W.M. *Tracking environmental change using lake sediments: terrestrial, algal, and siliceous indicators*. Dordrecht: Springer Netherlands. pp. 155-202.

Bayramova, E., Petrova, D., Marchenkov, A., Morozov, A., Galachyants, Y., Zakharova, Y., Bedoshvili, Y. & Likhoshway, Y. 2024. Differential expression of stress adaptation genes in a diatom *Ulnaria acus* under different culture conditions. *International Journal of Molecular Sciences*, 25(4):2314.

Bengio, Y., Courville, A., & Vincent, P. 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798-1828.

Béres, V., Stenger-Kovács, C., Buczkó, K., Padisák, J., Selmeczy, G. B., Lengyel, E., & Tapolczai, K. 2023. Ecosystem services provided by freshwater and marine diatoms. *Hydrobiologia*, 850(12):2707-2733.

BGCMA (Breede-Gouritz Catchment Management Agency). 2023. *Annual report 2022/23.* Cape Town: BGCMA.

Binta Islam, S., Valles, D., Hibbitts, T. J., Ryberg, W. A., Walkup, D. K., & Forstner, M. R. 2023. Animal species recognition with deep convolutional neural networks from ecological camera trap images. *Animals*, 13(9):1526.

Bishop, I., Boldrini, A., Clymans, W., Hall, C., Moorhouse, H., Parkinson, S., Scott-Somme, K., Thornhill, I. & Loiselle, S. 2025. FreshWater Watch: Investigating the health of freshwater ecosystems, from the bottom Up. *Citizen Science: Theory and Practice*, 10(1).

Bleckmann, C. A., Rabe, B., Edgmon, S. J., & Fillingame, D. 1995. Aquatic toxicity variability for fresh-and saltwater species in refinery wastewater effluent. *Environmental Toxicology and Chemistry*, 14(7):1219-1223.

Bonada, N., Prat, N., Resh, V.H. & Statzner, B. 2006. Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. *Annual Review of Entomology*, 51(1):495-523.

Borowiec, M.L., Dikow, R.B., Frandsen, P.B., McKeeken, A., Valentini, G. & White, A.E. 2022. Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, 13(8):640-1660.

Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, 6(10):1210-1218.

Brown, D.S., Marsden, G. & Loudon, M. 2011. WATER Alert! Disseminating drinking water quality information to South Africans. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 1915-1920.

Cade, B. S., & Noon, B. R. 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1(8):412-420.

Carroll, A. S., & Curtis, C. J. 2021. Increasing nutrient influx trends and remediation options at Hartbeespoort Dam, South Africa: a mass-balance approach. *Water SA*, 47(2):210-220.

CEMAGREF. 1982. Etude des méthodes biologiques quantitatives d'appréciation de la qualité des eaux. Rapport Division Qualit6 des Eaux Lyon - Agence Financière de Bassin Rhône-Méditerrané- Corse. Pierre-Bénite.

Chen, X. H., Li, Y. Y., Zhang, H., Liu, J. L., Xie, Z. X., Lin, L., & Wang, D. Z. 2018. Quantitative proteomics reveals common and specific responses of a marine diatom *Thalassiosira pseudonana* to different macronutrient deficiencies. *Frontiers in Microbiology*, 9:2761.

Chen, X., Wang, Y., Wei, H., & Zhang, J. 2024. Nitric acid rain decreases soil bacterial diversity and alters bacterial community structure in farmland soils. *Agronomy*, 14(5):971.

Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. 2020. A review of the artificial neural network models for water quality prediction. *Applied Sciences*, 10(17):5776.

Ching, C.Y., Lee, H.Y., Toriman, M.E., Abdullah, M., Yatim, B.B. 2015. Effect of the big flood events on the water quality of the Muar River, Malaysia. *Sustainable Water Resources Management*, 1(2):97-110.

Chonova, T., Kurmayer, R., Rimet, F., Labanowski, J., Vasselon, V., Keck, F., Illmer, P. & Bouchez, A., 2019. Benthic diatom communities in an alpine river impacted by waste water treatment effluents as revealed using DNA metabarcoding. *Frontiers in Microbiology*, *10*,:653.

Choudhury, A. R., Singh, N., Veeraraghavan, A., Gupta, A., Palani, S. G., Mehdizadeh, M., Omidi, A., & Al-Taey, D. K. A. 2023. Ascertaining and optimizing the water footprint and sludge management practice in steel industries. *Water*, 15(12): 2177.

Christin, S., Hervet, É. & Lecomte, N. 2019. *Applications for deep learning in ecology. Methods in Ecology and Evolution*, 10(10):1632–1644.

Chutter, F. M. 1994. *Research on the rapid biological assessment of water-quality impacts in streams and rivers*. Progress Report. WRC Report No. 422/1/98. Pretoria: Water Research Commission.

Chutter, F.M. 1972. An empirical biotic index of the quality of water in South African streams and rivers. *Water research*, 6(1):19-30.

Claassen, M. 2013. Integrated water resource management in South Africa. *International Journal of Water Governance*, 1(3-4):323-338.

Codd, G. A., Ward, C. J., & Bell, S. G. 1997. Cyanobacterial toxins: occurrence, modes of action, health effects and exposure routes. In *Applied Toxicology: Approaches Through Basic Science: Proceedings of the 1996 EUROTOX Congress Meeting Held in Alicante, Spain, September 22-25, 1996* (pp. 399-410). Berlin, Heidelberg: Springer Berlin Heidelberg.

Compas, E.D. & Wade, S., 2018. Testing the waters: a demonstration of a novel water quality mapping system for citizen science groups. *Citizen Science: Theory and Practice*, 3(2):1-11.

Coste, M. & Ayphassorho, H. 1991.Étude de la qualité des eaux du Bassin Artois-Picardie à l'aide des communautés de diatomées benthiques (application des indices diatomiques)

Cristóbal, E., Ayuso, S. V., Justel, A., & Toro, M. 2014. Robust optima and tolerance ranges of biological indicators: a new method to identify sentinels of global warming. *Ecological Research*, 29(1):55-68.

Dabrowski, J. M., Dabrowski, J., Hill, L., Macmillan, P., & Oberholster, P. J. 2015. Fate, transport and effects of pollutants originating from acid mine drainage in the Olifants River, South Africa. *River Research and Applications*, 31(10):1354-1364.

Dallas, H.F. & Day, J.A. 2004. The effect of water quality variables on aquatic ecosystems: A review. WRC Report No. TT 224/04. Pretoria: Water Research Commission.

Dallas, H.F. 2007. The effect of biotope-specific sampling for aquatic macroinvertebrates on reference site classification and the identification of environmental predictors in Mpumalanga, South Africa. *African Journal of Aquatic Science*, 32(2):165-173.

Dalu, T. & Froneman, P.W. 2016. Diatom-based water quality monitoring in southern Africa: challenges and future prospects. *Water SA*, 42(4):551-559.

Dalu, T., Cuthbert, R. N., Weyl, O. L., & Wasserman, R. J. 2022. Community structure and environmental factors affecting diatom abundance and diversity in a Mediterranean climate river system. *Science of The Total Environment*, 810:152366.

Dalu, T., Froneman, P.W., Chari, L.D. & Richoux, N.B. 2014. Colonisation and community structure of benthic diatoms on artificial substrates following a major flood event: A case of the Kowie River (Eastern Cape, South Africa). *Water SA*, 40(3):471-480.

Davenport, R. J., Satchell, M., & Shaw-Taylor, L. M. W. 2018. Cholera as a 'sanitary test' of British cities, 1831–1866, The History of the Family.

de Loë, R., 2009. Sharing the waters of the Red River basin: a review of options for transboundary water governance. Prepared for International Red Rivers Board, International Joint Commission. Guelph, ON: Rob de Loë Consulting Services.

De Villiers, S. D., & Thiart, C. 2007. The nutrient status of South African rivers: concentrations, trends and fluxes from the 1970s to 2005. *South African Journal of Science*, 103(7-8):343-349.

Death, R. G., Fuller, I. C., & Macklin, M. G. 2015. Resetting the river template: The potential for climate-related extreme floods to transform river geomorphology and ecology. *Freshwater Biology*, 60(12):2477-2496.

Dell'Aquila, G., Zauner, S., Heimerl, T., Kahnt, J., Samel-Gondesen, V., Runge, S., Hempel, F. & Maier, U.G. 2020. Mobilization and cellular distribution of phosphate in the diatom *Phaeodactylum tricornutum*. *Frontiers in Plant Science*, 11:579.

Deming, D. 2019. The aqueducts and water supply of ancient Rome. *Ground Water*, 58(1):152–161.

Descy, J. P. 1979. A new approach to water quality estimation using diatoms. *Nova Hedwigia Beiheft*, *64*:305-323.

Descy, J. P., & Coste, M. 1991. A test of methods for assessing water quality based on diatoms. *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen*, 24(4):2112-2116.

Di Cecco, G.J., Barve, V., Belitz, M.W., Stucky, B.J., Guralnick, R.P. & Hurlbert, A.H. 2021. Observing the observers: How participants contribute data to iNaturalist and implications for biodiversity science. *BioScience*, 71(11):1179-1188.

Díaz, S.M., Settele, J., Brondízio, E., Ngo, H., Guèze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K., Butchart, S. & Chan, K. 2019. The global assessment report on biodiversity and ecosystem services: Summary for policy makers.

Dickens, C.W. & Graham, P.M., 2002. The South African Scoring System (SASS) version 5 rapid bioassessment method for rivers. *African Journal of Aquatic Science*, 27(1):1-10.

Dieppois, B., Rouault, M. & New, M. 2015. The impact of ENSO on Southern African rainfall in CMIP5 ocean atmosphere coupled climate models. *Climate Dynamics*, 45:2425-2442.

Donato-R, J. C., Mendivelso, H. A., Pedraza-Garzón, E. L., & Sabater, S. 2022. Drivers of the diversity of diatoms in an oligotrophic Andean stream. *International Journal of Limnology*, 58:2.

Du Preez, G. C., Wepener, V., Fourie, H., & Daneel, M. S. 2018. Irrigation water quality and the threat it poses to crop production: evaluating the status of the Crocodile (West) and Marico catchments, South Africa. *Environmental Monitoring and Assessment*, 190(3):127.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.H., Soto, D., Stiassny, M.L. & Sullivan, C.A. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews*, 81(2):163-182.

Durkin, C. A., Koester, J. A., Bender, S. J., & Armbrust, E. V. 2016. The evolution of silicon transporters in diatoms. *Journal of Phycology*, 52(5):716-731.

DWA (Department of Water Affairs). 2011. Directorate Water Resource Classification. Classification of significant water resources in the olifants water management area (WMA 4): Report on Socio-economic evaluation and the decision-analysis framework. RDM/WMA04/00/CON/CLA/0411.

DWA (Department of Water Affairs). 2013. *National Water Resource Strategy: Water for an Equitable and Sustainable Future (Second Edition).* Pretoria: Department of Water Affairs.

DWAF (Department of Water Affairs and Forestry). 2004. Internal Strategic Perspective: Usutu to Mhlathuze Water Management Area. Report No. PB WMA 06/000/00/0304.

DWAF (Department of Water Affairs and Forestry). 2004. National Water Resource Strategy (NWRS), First Edition. Pretoria: South Africa.

DWAF (Department of Water Affairs and Forestry). 2007. National Water Resource Strategy (NWRS), Second Edition. Pretoria: South Africa.

DWAF (Department of Water Affairs and Forestry).1996a. South African Water Quality Guidelines. Vol 2: Recreational Water Use.

DWAF (Department of Water Affairs and Forestry).1996b. South African Water Quality Guidelines. Vol 7: Aquatic Ecosystems.

DWAF Department of Water Affairs and Forestry. 1986. Management of the water resources of the Republic of South Africa. Government printers: Pretoria.

DWS (Department of Water and Sanitation). 2014. Development of reconciliation strategies for large bulk water supply systems: Orange River. Report No. P RSA D000/00/18312/11.

DWS (Department of Water and Sanitation). 2016a. Determination of Resource Quality Objectives in the Mokolo, Matlabas, Crocodile West and Marico Catchments in the Limpopo North West Water Management Area (WMA01): Inception Report. Report No: RDM/WMA01/00/CON/RQO/0116.

DWS (Department of Water and Sanitation). 2016b. Water Quality Management Policies and Strategies for South Africa: Water Quality and Water Quality Management Challenges in South Africa. Edition DWS Report No: 000/00/21715/5. Pretoria: South Africa.

DWS (Department of Water and Sanitation). 2020. SDG 6.6.1 – 2020 Data Drive: RSA Response to UNEP (Final). Pretoria.

DWS (Department of Water and Sanitation). 2022a. River EcoStatus Monitoring Programme: State of River Report 2020-2021. Pretoria: Resource Quality Information Services.

DWS (Department of Water and Sanitation). 2022b. Business Case for the Vaal-Orange Catchment Management Agency.

DWS (Department of Water and Sanitation). 2022c. Investigation of groundwater and surface water interaction for the protection of water resources in the lower Vaal catchment: Report no: RDM/WMA05/00/GWSW/0423.

DWS (Department of Water and Sanitation). 2022d. *Data Management Strategy for Water and Sanitation in the Republic of South Africa, 1st Edition.* Pretoria: DWS.

DWS (Department of Water and Sanitation). 2023. *National Water Resource Strategy – 3rd Edition (NWRS-3).* Government Gazette, No. 47133. Pretoria: Government Printer.

Edokpayi, J. N., Rogawski, E. T., Kahler, D. M., Hill, C. L., Reynolds, C., Nyathi, E., Smith, J. A., Odiyo, J. O., Samie, A., Bessong, P., & Dillingham, R. 2018. Challenges to Sustainable Safe Drinking Water: A Case Study of Water Quality and Use across Seasons in Rural Communities in Limpopo Province, South Africa. *Water*, 10(2):159.

Ehrlich, P. R., & Mooney, H. A. 1983. Extinction, substitution, and ecosystem services. *BioScience*, *33*(4):248-254.

Eichinski, P., Alexander, C., Roe, P., Parsons, S., & Fuller, S. 2022. A convolutional neural network bird species recognizer built from little data by iteratively training, detecting, and labeling. *Frontiers in Ecology and Evolution*, 10:810330.

Erasmus, M. 2021. Evaluation of a diatom-based index for acid mine drainage impacted permanent pans in the Mpumalanga Highveld region. Potchefstroom: North-West University. (Dissertation - MSc).

Erasmus, J. H., Truter, M., Smit, N. J., Nachev, M., Sures, B., & Wepener, V. 2024. Element contamination of the Orange-Vaal River basin, South Africa: a one health approach. *Environmental Science and Pollution Research*, 31(20):29886-29901.

Erasmus, D. & Bega, S. 2025. School sampling tests find storage tanks contain unsafe water. The Green Guardian. https://mg.co.za/the-green-guardian/2025-06-25-school-sampling-tests-find-storage-tanks-contain-unsafe-water/?utm_source=chatgpt.com Date of access 21 Oct. 2025.

FAO Food and Agriculture Organization of the United Nations. 2018. The impact of disasters and crises on agriculture and food security. *Report*.

Fasihi, S., Lim, W. Z., Wu, W., & Proverbs, D. 2021. Systematic review of flood and drought literature based on science mapping and content analysis. *Water*, 13(19):2788. https://doi.org/10.3390/w13192788.

Feio, M. J., Viana-Ferreira, C., & Costa, C. 2014. Testing a multiple machine learning tool (HYDRA) for the bioassessment of fresh waters. *Freshwater Science*, 33(4):1286-1296.

Feio, M.J., Hughes, R.M., Callisto, M., Nichols, S.J., Odume, O.N., Quintella, B.R., Kuemmerlen, M., Aguiar, F.C., Almeida, S.F.P., Alonso-EguíaLis, P., Arimoro, F.O., Dyer, F.J., Harding, J.S., Jang, S., Kaufmann, P.R., Lee, S., Li, J., Macedo, D.R., Mendes, A., Mercado-Silva, N., Monk, W., Nakamura, K., Ndiritu, G.G., Ogden, R., Peat, M., Reynoldson, T.B., Rios-Touma, B., Segurado, P. & Yates, A.G. 2021. The biological assessment and rehabilitation of the world's rivers: An overview. *Water*, 13(3):371.

Fisheries Preservation Association. (1868). *On the pollution of the rivers of the kingdom: Evidence of the magnitude of the evil and the need for immediate legislative action.* London: Fisheries Preservation Association.

Fragoso, G. M., Poulton, A. J., Yashayaev, I. M., Head, E. J., Johnsen, G., & Purdie, D. A. 2018. Diatom biogeography from the Labrador Sea revealed through a trait-based approach. *Frontiers in Marine Science*, 5:297.

Freeman, M.C., Pringle, C.M. & Jackson, C.R., 2007. Hydrologic connectivity and the contribution of stream headwaters to ecological integrity at regional scales. *Journal of the American Water Resources Association*, 43(1):5-14.

Garg, R., & Singh, S. K. 2022. Treatment technologies for sustainable management of wastewater from iron and steel industry—a review. *Environmental Science and Pollution Research*, 29(50):75203-75222.

GCIS (Government Communication and Information System). 2022. *South Africa Yearbook 2021/22: Water and Sanitation*. Pretoria: GCIS.

Gerhardt, A. 2002. Bioindicator species and their use in biomonitoring. *Environmental Monitoring*, (1):77-123.

Gillwald, A., Mothobi, O. & Rademan, B. 2018. *The state of ICT in South Africa.* Policy Paper No. 5, Series 5: After access – Assessing digital inequality in Africa. Cape Town: Research ICT Africa.

Glevitzky, M., Corcheş, M. T., & Popa, D. M. 2025. Assessing pollution and diatom-based bioindicators in the Arieș River, Romania. *Environments*, 12(11):398.

GMSA (Global System for Mobile Communications Association). 2024. Driving digital transformation of the economy in South Africa opportunities, policy reforms and the role of mobile.

Godhe, A., & Rynearson, T. 2017. The role of intraspecific variation in the ecological and evolutionary success of diatoms in changing environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1728):20160399.

Goodfellow, I., Bengio, Y. & Courville, A. 2016. *Deep Learning*. Cambridge, MA: MIT Press.

Göthe, E., Angeler, D. G., Gottschalk, S., Löfgren, S., & Sandin, L. 2013. The influence of environmental, biotic and spatial factors on diatom metacommunity structure in Swedish headwater streams. *PloS one*, 8(8):e72237.

Graham, M. & Taylor, J., 2018. Development of citizen science water resource monitoring tools and communities of practice for South Africa, Africa and the world. WRC Report No. TT763/18. Pretoria: Water Research Commission.

Graham, M., Ross-Gillespie, V.& Taylor, J., 2016. A revised adopt-a-river programme: Stakeholder input on the institutional and financial frameworks with a focus on an implementation strategy: WRC Report No. KV354/16. Pretoria: Water Research Commission.

Graham, P. M., Pattinson, N. B., & Still, D. 2025. The state of wastewater management in South Africa: data gaps, missing wastewater, and Green Drop reporting. *Water SA*, 51(2):78-89.

Graham, P.M., Dickens, C.W. and Taylor, R.J., 2004. miniSASS - A novel technique for community participation in river health monitoring and management. *African Journal of Aquatic Science*, 29(1):25-35.

Graham, P.M., Pattinson, N.B., Russell, C. & Taylor, J. 2024. The value of citizen science for a just and sustainable water future. *South African Journal of Science*, 120(9/10).

Griffin, N. J. 2017. The rise and fall of dissolved phosphate in South African rivers. *South African Journal of Science*, 113(11-12):1-7.

Griffin, N.J., Palmer, C.G. Scherman, P.A. 2014. Critical analysis of environmental water quality in South Africa: Historic and current trends. WRC Report No. 2184/1/14. Pretoria: Water Research Commission.

Gründemann, G.J., Werner, M. & Veldkamp, T.I. 2018. The potential of global reanalysis datasets in identifying flood events in Southern Africa. *Hydrology and Earth System Sciences*, 22(9):4667-4683.

Guisan, A., & Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3):147-186.

GWP. 2000. Integrated Water Resources Management., TAC Background Paper No. 4. *Global Water Partnership Technical Advisory Committee, Stockholm*.

Halliday, S. 1999. *The Great Stink of London: Sir Joseph Bazalgette and the Cleansing of the Victorian Capital*. Stroud: Sutton Publishing.

Hanlon, W. W. 2020. Coal smoke, city growth, and the costs of the industrial revolution. *The Economic Journal*, 130(626):462-488.

Harding, W. R., & Taylor, J. C. 2011. The South African Diatom Index (SADI): a preliminary index for indicating water quality in rivers and streams in southern Africa: Report to the Water Research Commission. WRC Report No. 1707/1/11. Pretoria: Water Research Commission.

Harrison, T.D. & Whitfield, A.K., 2006. Application of a multimetric fish index to assess the environmental condition of South African estuaries. *Estuaries and Coasts*, 29(6):1108-1120.

Hawkes, H. A. 1997. Origin and development of the Biological Monitoring Working Party (BMWP) score system. *Water Research*, 32(3):964–968.

Hawkins, C.P., Norris, R.H., Hogue, J.N. & Feminella, J.W. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications*, 10(5):1456-1477.

Hendriks, H. & Rossouw, J.N., 2009. Adopt-a-River programme phase II: Development of an implementation plan. *Water Resource Quality Situation Assessment*.

Herbig, F. J. 2019. Talking dirty-effluent and sewage irreverence in South Africa: A conservation crime perspective. *Cogent Social Sciences*, 5(1):1701359.

Higgins, M. J., Crawford, S. A., Mulvaney, P., & Wetherbee, R. 2002. Characterization of the adhesive mucilages secreted by live diatom cells using atomic force microscopy. *Protist*, 153(1):25-38.

Hoell, A., Funk, C., Zinke, J. & Harrison, L. 2016. Modulation of the Southern Africa precipitation response to the El Niño Southern Oscillation by the subtropical Indian Ocean dipole. *Climate dynamics*, 48:2529-2540.

Høye, T.T., Dyrmann, M., Kjær, C., Nielsen, J., Bruus, M., Mielec, C.L., Vesterdal, M.S., Bjerge, K., Madsen, S.A., Jeppesen, M.R. & Melvad, C. 2022. Accurate image-based identification of macroinvertebrate specimens using deep learning—How much training data is needed? *PeerJ*, 10.

Hu, R., Xie, J., Wu, S., Yang, C., & Yang, D. 2020. Study of Toxicity Assessment of Heavy Metals from Steel Slag and Its Asphalt Mixture. *Materials*, 13(12):2768.

Hudon, C., Legendre, P. 1987. The ecological implications of growth forms in epibenthic diatoms. *Journal of Phycology,* 25:434-441.

Humphries, M. S., McCarthy, T. S., & Pillay, L. 2017. Attenuation of pollution arising from acid mine drainage by a natural wetland on the Witwatersrand. *South African Journal of Science*, 113(1-2):1-9.

iNaturalist. 2024. *Behind the power of iNaturalist*. https://www.inaturalist.org/posts/23833-behind-the-power-of-inaturalist?utm_source=chatgpt.com Date of access: 23 Feb. 2025.

IUCMA (Inkomati-Usuthu Catchment Management Agency). 2020. *Annual Performance Plan 2020/21.* Mbombela: IUCMA.

IUCMA (Inkomati-Usuthu Catchment Management Agency). 2023. *Groundwater Accounting for Inkomati-Usuthu WMA.* Mbombela: IUCMA.

IUCMA (Inkomati-Usuthu Catchment Management Agency). 2024. *Annual performance plan 2024/25.* Mbombela: IUCMA.

Jaganyi, J., Salagae, M., Matiwane. 2009. Integrating floodplain livelihoods into a diverse rural economy by enhancing co-operative management: A case study of the Pongolo floodplain system, South Africa. WRC Report No. 1299/1/08. Pretoria: Water Research Commission.

Jamil, T., Kruk, C., & ter Braak, C. J. 2014. A unimodal species response model relating traits to environment with application to phytoplankton communities. *PloS One*, 9(5):e97583.

Jiyane, P, Hoorzook, K.B. & van Rensburg, N.J. 2025. The state of wastewater treatment plants and how it is affecting the effluent discharge quality within the water bodies of Gauteng. in *Nature-Based Technologies for Wastewater Treatment and Bioenergy Production.* IWA Publishing, pp. 139-152.

Johnson, R.K., Wiederholm, T. & Rosenberg, D.M. 1993. Freshwater biomonitoring using individual organisms, populations, and species assemblages of benthic macroinvertebrates. In: Rosenberg, D.M. & Resh, V.H., eds. *Freshwater Biomonitoring and Benthic Macroinvertebrates*. Chapman & Hall. pp. 40-158.

Joutsijoki, H., Meissner, K., Gabbouj, M., Kiranyaz, S., Raitoharju, J., Ärje, J., Kärkkäinen, S., Tirronen, V., Turpeinen, T. & Juhola, M. 2014. Evaluating the performance of artificial neural networks for the classification of freshwater benthic macroinvertebrates. *Ecological Informatics*, 20:1-12.

Jwaideh, M. A., Sutanudjaja, E. H., & Dalin, C. 2022. Global impacts of nitrogen and phosphorus fertiliser use for major crops on aquatic biodiversity. *The International Journal of Life Cycle Assessment*, 27(8):1058-1080.

Kahlert, M., Fölster, J., & Tapolczai, K. 2023. No lukewarm diatom communities - the response of freshwater benthic diatoms to phosphorus in streams as basis for a new phosphorus diatom index (PDI_{SE}). *Environmental Monitoring and Assessment*, *195*(7):807.

Kang, L. K., & Rynearson, T. A. 2019. Identification and expression analyses of the nitrate transporter gene (NRT 2) family among Skeletonema species (Bacillariophyceae). *Journal of phycology*, 55(5):1115-1125.

Kaushal, S.S., Likens, G.E., Pace, M.L., Reimer, J.E., Maas, C.M., Galella, J.G., Utz, R.M., Duan, S., Kryger, J.R., Yaculak, A.M. and Boger, W.L. 2021. Freshwater salinization syndrome: from emerging global problem to managing risks. *Biogeochemistry*, 154(2):255-292.

Keck, F., & Kahlert, M. 2019. Community phylogenetic structure reveals the imprint of dispersal-related dynamics and environmental filtering by nutrient availability in freshwater diatoms. *Scientific Reports*, 9(1):11590.

Kelly, M.G. & Whitton, B.A. 1995. The trophic diatom index: A new index for monitoring eutrophication in rivers. *Journal of Applied Phycology*, 7:433-444.

Kleynhans, C.J. 2007. Module D: Fish Response Assessment Index in River EcoClassification: Manual for EcoStatus determination (version 2) Joint Water Research Commission and Department of Water Affairs and Forestry report.

Kleynhans, C.J., Louw, M.D. & Graham, M, 2008. Module G: EcoClassification and EcoStatus determination in River EcoClassification: Index of Habitat Integrity (Section 1, Technical manual) Joint Water Research Commission and Department of Water Affairs and Forestry report. WRC Report No. TT 377-08.

Kleynhans, C.J., Mackenzie, J., Louw, M.D. 2007. Module F: Riparian Vegetation Response Assessment Index in River EcoClassification: Manual for EcoStatus Determination (version 2). Joint Water Research Commission and Department of Water Affairs and Forestry report.

Kloos, J., & Tsegai, D. W. 2009. Preferences for domestic water services in the Middle Olifants sub-basin of South Africa, ZEF- Discussion Papers On Development PolicyNo. 131, Center for Development Research, Bonn, pp.23.

Kock, A. 2017. Diatom diversity and response to water quality within the Makuleke Wetlands and Lake Sibaya. Potchefstroom: North-West University. (Dissertation - MSc).

Koen, R.C.J., Koen, F.J., Pattinson, N.B., Dickens, C.W. & Graham, P.M. 2023. *Digitally improving the identification of aquatic macroinvertebrates for indices used in biomonitoring*. International Water Management Institute (IWMI). CGIAR Initiative on Digital Innovation.

Kolkwitz, R. & Marsson, M. 1908. *Ökologie der pflanzlichen Saprobien*. Berichte der Deutschen Botanischen Gesellschaft, 26A: 505–519.

Kosmala, M., Wiggins, A., Swanson, A. & Simmons, B. 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14(10):551-560.

Lake, P.S., 2000. Disturbance, patchiness, and diversity in streams. *Journal of the North American Benthological society*, 19(4):573-592.

Lam, Q.D., Rötter, R. P., Rapholo, E., Ayisi, K., Nelson, W. C. D., Odhiambo, J., & Foord, S. 2023. Modelling maize yield impacts of improved water and fertilizer management in southern Africa

using cropping system model coupled to an agro-hydrological model at field and catchment scale. *The Journal of Agricultural Science*, 161(3):356–372.

Lange-Bertalot, H. 1979. Pollution tolerance of diatoms as a criterion for water quality estimation. *Nova Hedwigia Beiheft*, *64*:285-304.

Lecointe, C., Coste, M. & Prygiel, J. 1993. “Omnidia”: software for taxonomy, calculation of diatom indices and inventories management. *Hydrobiologia*, 269(1):509-513.

LeCun, Y., Bengio, Y., & Hinton, G. 2015. Deep learning. *Nature*, 521(7553):436-444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 2002. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278-2324.

Lek, S. & Guégan, J. F. 2000. Artificial neuronal networks: Application to ecology and evolution. In: *Artificial neuronal networks: Application to ecology and evolution*. Eds: Lek, S. & Guégan, J. F. Berlin: Springer.

Lek, S. & Guégan, J.F. 1999. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, 120(2-3):65-73.

Lenders, H. R. 2017. Fish and fisheries in the Lower Rhine 1550–1950: A historical-ecological perspective. *Journal of Environmental Management*, 202:403-411.

Lenoir, A. & Coste, M. 1996. *Development of a practical diatom index of overall water quality applicable to the French National Water Board network.* In Whitton, B. A. & Rott, E. Eds: *Use of Algae for Monitoring Rivers II*. Austria: Universität Innsbruck. pp. 29-43.

Li, H.R., Xiang, H.M., Zhong, J.W., Ren, X.Q., Wei, H., Zhang, J.E., Xu, Q.Y., & Zhao, B.L. 2020. Acid rain increases impact of rice blast on crop health via inhibition of resistance enzymes. *Plants*, 9(7):881.

Li, Z., Zhang, Y., Li, W., Irwin, A. J., & Finkel, Z. V. 2023. Common environmental stress responses in a model marine diatom. *New Phytologist*, 240(1):272-284.

Lombaard, J., Pieterse, H. S., Rossouw, J. R., & Nditwani, T. 2015. Project Name: Limpopo Water Management Area North Reconciliation Strategy – Literature Review.

López-Guillén, E., Herrera, I., Bensid, B., Gómez-Bellver, C., Ibáñez, N., Jiménez-Mejías, P., Mairal, M., Mena-García, L., Nualart, N., Utjés-Mascó, M. & López-Pujol, J. 2024. Strengths and challenges of using iNaturalist in plant research with focus on data quality. *Diversity*, 16(1):42.

Lu, X., Liu, Y., & Fan, Y. 2020. Diatom taxonomic composition as a biological indicator of the ecological health and status of a river basin under agricultural influence. *Water*, 12(7):2067.

Lukhele, T., & Msagati, T. A. M. 2024. Eutrophication of inland surface waters in South Africa: An overview. *International Journal of Environmental Research*, 18(2):27.

Luna, S., Gold, M., Albert, A., Ceccaroni, L., Claramunt, B., Danylo, O., Haklay, M., Kottmann, R., Kyba, C., Piera, J. & Radicchi, A. (2018). Developing mobile applications for environmental and biodiversity citizen science: considerations and recommendations. In: Multimedia tools and applications for environmental & biodiversity informatics. Cham: Springer International Publishing. pp. 9–30.

Lwin, K., Hashimoto, M. & Murayama, Y. 2014. Real-time geospatial data collection and visualization with smartphone. *Journal of Geographic Information System*, 6(2):99-108.

Malan, H. L., & Day, J. A. 2005. Wetland water quality and the Ecological Reserve. WRC Report No. 1311/1/05. Pretoria: Water Research Commission.

Malan, H., Day, J., Ramjukadh, C. & Olivier, N., 2015. Trajectories of changes in wetlands of the Fynbos Biome from the late 1980s to 2014. WRC Report No. 2183/1/14. Pretoria: South Africa.

Malthus, T. J., Ohmsen, R., & van der Woerd, H. 2020. An Evaluation of Citizen Science Smartphone Apps for Inland Water Quality Assessment. *Remote Sensing*, 12(10):1578.

Mangadze, T., Dalu, T., & Froneman, P. (2019). Biological monitoring in southern Africa: A review of the current status, challenges and future prospects. *The Science of The Total Environment*, 648:1492–1499.

Mararakanye, N., Le Roux, J. J., & Franke, A. C. 2022. Long-term water quality assessments under changing land use in a large semi-arid catchment in South Africa. *Science of The Total Environment*, 818:151670.

Masouras, A., Karaouzas, I., Dimitriou, E., Tsirtsis, G., & Smeti, E. 2021. Benthic diatoms in river biomonitoring—present and future perspectives within the water framework directive. *Water*, 13(4):478.

Matsui, H., Harada, H., Maeda, K., Sugiyama, T., Fukuchi, Y., Kimura, N., Nawaly, H., Tsuji, Y. and Matsuda, Y. 2024. Coordinated phosphate uptake by extracellular alkaline phosphatase and solute carrier transporters in marine diatoms. *New Phytologist*, 241(3):1210-1221.

Matta, S., Taylor, J. & Jewitt, G. 2020. *Opportunities for Citizen Science to Improve Catchment Water Quality.* WRC-EI Deliverable 5. Pietermaritzburg: Centre for Water Resources Research, University of KwaZulu-Natal.

McCarthy, T. S. 2011. The impact of acid mine drainage in South Africa. *South African Journal of Science*, 107(5):1-7.

Mickelsson, M., Thifhulufhelwi, R., Mvulane, P., Brownell, F., Russell, C. & Lotz-Sisitka, H. 2024. Bringing river health into being with citizen science: River commons co-learning and practice. *South African Journal of Science*, 120(9/10).

Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-being: Synthesis. Washington, DC.

Milosavljević, A., Milošević, Đ., & Predić, B. 2021. Species identification for aquatic biomonitoring using deep residual CNN and transfer learning. *Facta Universitatis, Series: Automatic Control and Robotics*, 20(1):1-19.

Milošević, D., Milosavljević, A., Predić, B., Medeiros, A.S., Savić-Zdravković, D., Piperac, M.S., Kostić, T., Spasić, F. & Leese, F. 2020. Application of deep learning in aquatic bioassessment: Towards automated identification of non-biting midges. *Science of the Total Environment*, 711:135160.

Minsky, M. & Papert, S. 1969. *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.

Mitchell, J. G., Seuront, L., Doubell, M. J., Losic, D., Voelcker, N. H., Seymour, J., & Lal, R. 2013. The role of diatom nanostructures in biasing diffusion to improve uptake in a patchy nutrient environment. *PLoS One*, 8(5):e59548.

Mitchell, S. A., De Wit, M. P., Blignaut, J. N., & Crookes, D. 2014. Waste water treatment plants: The financing mechanisms associated with achieving green drop rating. WRC Report No. 2085/1/14. Pretoria: Water Research Commission.

Mnisi, N. 2020. *Water scarcity in South Africa: A result of physical or economic factors?* https://hsf.org.za/publications/hsf-briefs/water-scarcity-in-south-africa-a-result-of-physical-or-economic-factors. Date of access: 10 Nov. 2024.

Mnyango, S. S., Thwala, M., Truter, C., Xulu, N. G., Schoeman, Y., & Oberholster, P. J. 2025. Drivers and management of nutrient overload in dams: Insights from Roodeplaat Dam, South Africa. *Hydrology,* 12(3):2306-5338.

Mohammadi, L., Rahdar, A., Bazrafshan, E., Dahmardeh, H., Susan, M. A. B. H., & Kyzas, G. Z. 2020. Petroleum Hydrocarbon Removal from Wastewaters: A Review. *Processes*, 8(4):447.

Moodley, P., Juggath, N., Vemblanathan, Y., van Niekerk, M.A., Petersen, Y., Baloyi, M., Pulles, W., Love, D & Aspeling, B. 2021. Water conservation / water demand management in the South African mining industry: A compendium of best practices and case studies. WRC Report No. TT 846/20. Pretoria: Water Research Commission.

Mudaly, L., & van der Laan, M. 2020. Interactions between irrigated agriculture and surface water quality with a focus on phosphate and nitrate in the Middle Olifants catchment, South Africa. *Sustainability*, 12(11):4370.

Mulangaphuma, L., & Jovanovic, N. 2025. Investigation of water use and trends in South Africa: A case study for the Mzimvubu to Tsitsikamma Water Management Area 7 (WMA7). *Water*, 17(10):1522.

Munyai, R., Ogola, H. J. O., & Modise, D. M. (2021). Microbial community diversity dynamics in acid mine drainage and acid mine drainage-polluted soils: implication on mining water irrigation agricultural sustainability. *Frontiers in Sustainable Food Systems*, 5: 701870.

Musingafi, M. C. C., & Tom, T. 2014. Fresh water sources pollution: A human related threat to fresh water security in South Africa*. Journal of Public Policy & Governance*, 1(2):72–81.

Nakov, T., Ashworth, M., & Theriot, E. C. 2015. Comparative analysis of the interaction between habitat and growth form in diatoms. *The ISME Journal*, 9(1):246-255.

Nhesvure, B. 2020. Impacts of ENSO on coastal South African sea surface temperatures. Cape Town: University of Cape Town. (Dissertation - MSc).

Nicholson, S.E. & Selato, J.C. 2000. The influence of La Niña on African rainfall. *International Journal of Climatology*, 20:1761-1776.

NielsenIQ. 2018. *Smartphones shine in a difficult year for South Africa's consumer technology sector*. https://www.bizcommunity.com/Article/196/168/175151.html?utm_source=chatgpt.com Date of access: 15 Jan. 2024.

Nikolaou, A., & Katsanevakis, S. 2023. Marine extinctions and their drivers. *Regional Environmental Change*, 23(3):88.

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115(25):5716-5725.

Nqowana, T., Fogel, R., Bezerra, J.C.& Limson, J. 2024. Citizen science tools for engaged research: Water quality monitoring in remote communities. *South African Journal of Science*, 120(9/10).

NWA (National Water Act). 1998. Government Gazette, Vol. 398, No. 19182. Pretoria: Government Printer.

Oberholster, P. J., & Ashton, P. J. 2008. State of the nation report: An overview of the current status of water quality and eutrophication in South African rivers and reservoirs. *Parliamentary Grant Deliverable. Pretoria: Council for Scientific and Industrial Research (CSIR)*, 1273-1280.

Olden, J.D. & Jackson, D.A. 2002. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154(1-2):135-150.

Ollis, D.J., Dallas, H.F., Esler, K.J. & Boucher, C., 2006. Bioassessment of the ecological integrity of river ecosystems using aquatic macroinvertebrates: An overview with a focus on South Africa. *African Journal of Aquatic Science*, 31(2):205-227.

Otto, M. 2018. Diatom community composition and ecological gradients on selected rivers in the Eastern and Western Cape, South Africa. Bloemfontein: University of the Free State (Dissertation - PhD).

Owolabi, S.T., Belle, J.A. 2023. Investigating extreme hydrological risk impact on water quality; evidence from Buffalo catchment headwater, Eastern Cape, South Africa. *Environmental Science and Pollution Research*, 30:73425–73450.

Palmer, C. G., & Coleman, H. D. 2004. Applied aquatic ecotoxicology: sub-lethal methods, whole effluent testing and communication. WRC Report No. 1245/1/04. Pretoria: Water Research Commision.

Pantle, R., & Buck, H. 1955. Die biologische Uberwachung der Gewässer und die Darstellung der Ergebnisse.

Passy, S. 2007. Diatom ecological guilds display distinct and predictable behaviour along nutrient and disturbance gradients in running waters. *Aquatic Botany*, 86:171-178.

Patil, J. S., & Anil, A. C. 2005. Biofilm diatom community structure: influence of temporal and substratum variability. *Biofouling*, 21(3-4):189-206.

Patrick, R. 1977. Ecology of Freshwater Diatoms – Diatom Communities. In: Werner, D. ed. The biology of diatoms. Berkeley: University of California Press. pp. 284-326.

Pattinson, N.B., Russell, C., Taylor, J., Dickens, C.W., Koen, R.C.J., Koen, F.J. & Graham, P.M. 2023. *Digital innovation with miniSASS, a citizen science biomonitoring tool*. Colombo, Sri Lanka: International Water Management Institute (IWMI). CGIAR Initiative on Digital Innovation. 11p.

Pichler, M., & Hartig, F. 2023. Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, 14(4):994-1016.

Qamar, R. & Zardari, B.A. 2023. Artificial neural networks: An overview. *Mesopotamian Journal of Computer Science*, 2023:124-133.

Qiu, J., Shen, Z., & Xie, H. 2023. Drought impacts on hydrology and water quality under climate change. *Science of The Total Environment*, *858*: 159854.

Quetglas, A., Ordines, F. & Guijarro, B. 2011. The use of artificial neural networks (ANNs) in aquatic ecology. In: Hui, C. L. P. (ed.) *Artificial Neural Networks – Application*. London: IntechOpen.

Rawat, W., & Wang, Z. 2017. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9):2352-2449.

Reddick, J., Maherry, A., Roux, S., Fourie, M., Hugo, R., Horsfield, C., & Loser, N. 2018. Water impacts and externalities of coal power. *Cape Town, South Africa: Life After Coal/Impilo Ngaphandle Kwamalahle Campaign*.

Riato, L., Leira, M., Della Bella, V., & Oberholster, P. J. 2018. Development of a diatom-based multimetric index for acid mine drainage impacted depressional wetlands. *Science of The Total Environment*, 612:214-222.

Rimet, F., & Bouchez, A. 2011. Use of diatom life-forms and ecological guilds to assess pesticide contamination in rivers: Lotic mesocosm approaches. *Ecological Indicators*, 11(2):489–499.

Rimet, F., & Bouchez, A. 2012a. Biomonitoring river diatoms: Implications of taxonomic resolution. *Ecological Indicators*, 15:92–99.

Rimet, F., & Bouchez, A. 2012b. Life-forms, cell sizes and ecological guilds of diatoms in European rivers. *Knowledge and Management of Aquatic Ecosystems*, 406(1):1–12.

Rodrigues, F. A., & Joekes, I. 2011. Cement industry: sustainability, challenges and perspectives. *Environmental Chemistry Letters*, 9(2):151-166.

Rosenblatt, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review,* 65(6):386–408.

Rott, E., Hofmann, G., Pall, K., Pfister, P. & Pipp, E. 1997. *Indikationslisten für Aufwuchsalgen, Teil 1: Saprobielle Indikation.* Bundesministerium für Land- und Forstwirtschaft, Wien, 73 pp.

Rott, E., Pipp, E. & Pfister, P. 2003. Diatom methods developed for river quality assessment in Austria and a cross-check against numerical trophic indication methods used in Europe. *Algological Studies*, 110:91–115.

Rott, E., Pipp, E., Pfister, P., van Dam, H., Ortler, K., Binder, N. & Pall, K. 1999. *Indikationslisten für Aufwuchsalgen, Teil 2: Trophieindikation sowie geochemische Präferenz, taxonomische und toxikologische Anmerkungen.* Bundesministerium für Land- und Forstwirtschaft, Wien, 248 pp.

Round, F.E., Crawford, R.M. & Mann, D.G., 1990. *Diatoms: biology and morphology of the genera*. Cambridge University Press. UK.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986. Learning representations by back-propagating errors. *Nature*, 323: 533–536.

Sala, O. E., Stuart Chapin, F. I. I, Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R. & Wall, D. H. 2000. Global biodiversity scenarios for the year 2100. *Science*, *287*(5459):1770-1774.

Saleem, U., Torabi Haghighi, A., Klöve, B. & Oussalah, M. 2024. Citizen science applications for water quality monitoring. A review. *EGUsphere*, 2024:1-18.

Sandha, S.S., Srivastava, B. & Randhawa, S. 2017. The GangaWatch mobile app to enable usage of water data in every day decisions integrating historical and real-time sensing data.

Sarker, I. H. 2021. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6):1-20.

Schreiner, B. 2013. Viewpoint - Why has the South African National Water Act been so difficult to implement. *Water Alternatives*, 6(2):239-245.

Schuler, M. S., Cañedo-Argüelles, M., Hintz, W. D., Dyack, B., Birk, S., & Relyea, R. A. 2019. Regulations are needed to protect freshwater ecosystems from salinization. *Philosophical Transactions of the Royal Society B*, 374(1764):20180019.

Serna López, J. P., Fernández Mc Cann, D. S., Vélez Macías, F. D. J., & Aguirre Ramírez, N. J. 2020. An image processing method for recognition of four aquatic macroinvertebrates genera in freshwater environments in the Andean region of Colombia. *Environmental Monitoring and Assessment*, 192(10):617.

Shibabaw, T., Beyene, A., Awoke, A., Tirfie, M., Azage, M., & Triest, L. 2021. Diatom community structure in relation to environmental factors in human influenced rivers and streams in tropical Africa. *Plos one*, 16(2):e0246043.

Shikwambana, P., Taylor, J. C., Govender, D., & Botha, J. 2021. Diatom responses to river water quality in the Kruger National Park, South Africa. *Bothalia-African Biodiversity & Conservation*, 51(1):1-11.

Shin, Y., Kim, D., & Heo, T. Y. 2022. Determination of the habitat preferences of dominant epilithic diatoms using statistical models: a case study in the Han River, South Korea. *Water*, 14(6):956.

Sitoe, S., & Qwist-Hoffman, P. 2013. Limpopo River Basin Monograph LRBMS-81137945. Report No:8520.

Sládecek, V. 1973. System of water quality from the biological point of view. *Archiv fur Hydrobiologie (Ergebnisse der Limnologie)*, *7*:1-218.

Smol, J.P. & Stoermer, E.F. eds. 2010. *The diatoms: applications for the environmental and earth sciences*. 2nd ed. Cambridge: Cambridge university press.

Smol, J.P. 2008. *Pollution of lakes and rivers: a paleoenvironmental perspective*. 2nd ed. Malden, MA: Blackwell Publishing Ltd.

Solomon, K. J., Stevenson, R. J., Surratt, D., Whelan, K. R., Tobias, F. A., Johnson, K. M., & Gaiser, E. E. 2025. Robust species-optima estimates from non-uniformly sampled environmental gradients. *Journal of Paleolimnology*, 1-16.

Sosibo, N. Z., Muchaonyerwa, P., Visser, L., Barnard, A., Dube, E., & Tsilo, T. J. 2017. Soil fertility constraints and yield gaps of irrigation wheat in South Africa. *South African Journal of Science*, 113(1-2):1-9.

Steele, D. J., Franklin, D. J., & Underwood, G. J. 2014. Protection of cells from salinity stress by extracellular polymeric substances in diatom biofilms. *Biofouling*, 30(8):987-998.

Stefanakis, A. 2021. Promoting sustainability in the oil industry: The benefits of using constructed wetlands for oily wastewater treatment. *Sustainable Industrial Water Use: Perspectives, Incentives, and Tools*, *427*.

Stenger-Kovács, C., & B-Béres, V. 2024. Guilds for diatoms: History and future prospects. *Diatom Ecology: From Molecules to Metacommunities*, 233-279.

Stenger-Kovács, C., Béres, V.B., Buczkó, K., Tapolczai, K., Padisák, J., Selmeczy, G.B. & Lengyel, E. 2023. Diatom community response to inland water salinization: a review. *Hydrobiologia*, 850(20):627-4663.

Stevenson, R. J., Pan, Y., Manoylov, K. M., Parker, C. A., Larsen, D. P., & Herlihy, A. T. 2008. Development of diatom indicators of ecological conditions for streams of the western US. *Journal of the North American Benthological Society*, 27(4):1000-1016.

Steyn, N.R. 2022, September. Legislative responses to the challenge of insufficient data on water service delivery in South African cities. *Urban Forum* ,33(3):349-366).

Su, Y., Chen, J., Hu, J., Qian, C., Ma, J., Brynjolfsson, S., & Fu, W. 2024. Manipulation of ion/electron carrier genes in the model diatom Phaeodactylum tricornutum enables its growth under lethal acidic stress. *iScience*, 27(8).

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D. & Kelling, S. 2009. eBird: A citizen-based bird observation network in the biological sciences. *Biological conservation*, 142(10):2282-2292.

Szczepocka, E. & Żelazna-Wieczorek, J. 2018. Diatom biomonitoring–scientific foundations, commonly discussed issues and frequently made errors. *Oceanological and Hydrobiological Studies*, 47(3):313-325.

Szczepocka, E., Szulc, B., Szulc, K., Rakowska, B. & Żelazna-Wieczorek, J. 2014. Diatom indices in the biological assessment of the water quality based on the example of a small lowland river. *Oceanological and Hydrobiological Studies*, 43(3):265-273.

Tabasi, M., Siripanich, A., Khan, N.A., Auld, J. & Rashidi, T.H. 2024. GPS-supported smartphone app-based integrated travel diary and time-use data collection: challenges and lessons learned. *Transportation*, pp.1-26.

Talbot, C.J., Bennett, E.M., Cassell, K., Hanes, D.M., Minor, E.C.,Paerl, H., Raymond, P.A., Vargas, R., Vidon, P.G., Wollheim, W. & Xenopoulos, M.A. 2018. The impact of flooding on aquatic ecosystem services. *Biogeochemistry*, 141(3):439-461.

Taylor, J. C., & Cocquyt, C. 2015. Diatom research in southern and central Africa: Historical perspectives and current activities. *Mededelingen van de Zittingen van de Koninklijke Academie voor Overzeese Wetenschappen*, 61(4):593-610.

Taylor, J., Graham, M., Louw, A., Lepheana, A., Madikizela, B., Dickens, C., Chapman, D.V. & Warner, S. 2022. Social change innovations, citizen science, miniSASS and the SDGs. *Water Policy*, 24(5):708-717.

Taylor, J.C. 2004. The Application of diatom-based pollution indices in the Vaal Catchment. Potchefstroom: North-West University. (Dissertation - MSc).

Taylor, J.C. 2007. Towards the successful application of diatom-based biomonitoring in South Africa. Potchefstroom: North-West University. (Dissertation - PhD).

Taylor, J.C., de la Rey, P.A. & van Rensburg, L., 2005. Recommendations for the collection, preparation and enumeration of diatoms from riverine habitats for water quality monitoring in South Africa. *African Journal of Aquatic Science*, 30(1):65-75.

Taylor, J.C., Harding, W.R. & Archibald, C.G.M. 2007b. *A methods manual for the collection, preparation and analysis of diatom samples*. WRC Report No. TT281/07. Pretoria: Water research Commission.

Taylor, J.C., Harding, W.R. and Archibald, C.G.M. 2007c. *An illustrated guide to some common diatom species from South Africa*. WRC Report No. TT282/07. Pretoria: Water Research Commission.

Taylor, J.C., Prygiel, J., Vosloo, A., de la Rey, P.A. & van Rensburg, L. 2007a. Can diatom-based pollution indices be used for biomonitoring in South Africa? A case study of the Crocodile West and Marico water management area. *Hydrobiologia*, 592(1):455-464.

Teacher, A.G., Griffiths, D.J., Hodgson, D.J. & Inger, R. 2013. Smartphones in ecology and evolution: A guide for the app-rehensive. *Ecology and Evolution*, 3(16):5268-5278.

Ter Braak, C. J. 1987. *Unimodal models to relate species to environment*. Wageningen University and Research.

Ter Braak, C. J., & Juggins, S. 1993. Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia*, 269(1):485-502.

Ter Braak, C. J., & Looman, C. W. 1986. Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio*, 65(1):3-11.

Ter Braak, C. J., & Van Tongeren, O. F. R. 1995. *Data analysis in community and landscape ecology*. Cambridge: Cambridge Univ Press.

Ter Braak, C. J., & Verdonschot, P. F. 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic sciences*, 57(3):255-289.

Thienemann A, .1910. Das sammeln von puppenhauten der chironomiden. *Archiv Fur Hydrobiologie*, 6:213-214.

Thirion, C., 2007. Module E: Macro-Invertebrate response assessment index (MIRAI). River ecoclassification manual for ecostatus determination (Version 2): Joint Water Research Commission and Department of Water Affairs and Forestry report.

Thorslund, J., Bierkens, M. F., Oude Essink, G. H., Sutanudjaja, E. H., & van Vliet, M. T. 2021. Common irrigation drivers of freshwater salinisation in river basins worldwide. *Nature Communications*, 12(1):4232.

Tison-Rosebery, J., Boutry, S., Bertrin, V., Leboucher, T., & Morin, S. 2023. A new diatom-based multimetric index to assess lake ecological status. *Environmental Monitoring and Assessment*, 195(10):1202.

Tredoux, G. 2004. *Nitrate and associated hazard quantification and strategies for protecting rural water supplies*. WRC Report No. 1058/1/04. Pretoria: Water Research Commission.

Truong, M.X.A. & Van der Wal, R. 2024. Exploring the landscape of automated species identification apps: Development, promise, and user appraisal. *BioScience*, 74(9):601-613.

Valenti, S. 2024. 'Who can tell me what potable water means?' The assessment of water quality in debates over hydraulic infrastructure in nineteenth-century Italy. *The British Journal for the History of Science*, 58:101-116.

Van Vliet, M. & Zwolsman, G. 2008. Impact of summer droughts on the water quality of the Meuse River. *Journal of Hydrology*, 353(1):1-17.

Vanelslander, B., De Wever, A., Van Oostende, N., Kaewnuratchadasorn, P., Vanormelingen, P., Hendrickx, F., Sabbe, K. & Vyverman, W. 2009. Complementarity effects drive positive diversity effects on biomass production in experimental benthic diatom biofilms. *Journal of Ecology*, 97(5):1075-1082.

Várbíró, G., Borics, G., Novais, M.H., Morais, M.M., Rimet, F., Bouchez, A., Tapolczai, K., Bácsi, I., Usseglio-Polatera, P. & B-Béres, V. 2020. Environmental filtering and limiting similarity as main forces driving diatom community structure in Mediterranean and continental temporary and perennial streams. *Science of The Total Environment*, 741:140459.

Varjani, S., Joshi, R., Srivastava, V. K., Ngo, H. H., & Guo, W. 2020. Treatment of wastewater from petroleum industry: current practices and perspectives. *Environmental Science and Pollution Research*, 27(22):27172-27180.

Von Stosch, H.A. 1974. Pleurax, seine synthese und seine verwendung zur einbettung und darstellung det zellwande von diatomeen, peridineen und anderen algen, sowie fur eine neue methode zur electivfarbung von dinoflagellaten-penzern. *Archiv fur Protistenkunde*, 116:132-141.

Wang, X., & Tang, K. W. 2022. Chain or sphere? Perspectives on colony shapes and sizes in microalgae. *Journal of Plankton Research*, 44(4):521-527.

Watson, S. B. (2004). Aquatic taste and odor: a primary signal of drinking-water integrity. *Journal of Toxicology and Environmental Health, Part A*, 67(20-22):1779-1795.

Wilkinson, K., Palmer, C.G. & Odume, O.N., 2016. National aquatic ecosystem health monitoring in South Africa: progress and approaches. *Water SA*, 42(3):395–408.

Williams, W. D. 1999. Salinisation: A major threat to water resources in the arid and semi-arid regions of the world. *Lakes & Reservoirs: Research & Management*, 4(3-4):85-91.

Windisch, J., Gradwohl, A., Gilbert, B. M., Dos Santos, Q. M., Avenant-Oldewage, A., & Jirsa, F. 2025. Metals and metalloids in macrophytes and fish from an acid mine drainage-impacted river system in South Africa: aspects of bioindication and phytoremediation. *Environmental Science: Advances*, 4(9):1501-1513.

Wyrzykowski, R. 2023. Rural-urban digital divide still poses a challenge in South Africa. Open Signal. https://www.opensignal.com/2023/12/13/rural-urban-digital-divide-still-poses-a-challenge-in-south-africa?utm_source=chatgpt.com Date of access 22 Oct. 2025.

Yogeshwaran, M., Thacker, M., Wadmare, N., Vigneshwaran, A., Radhakrishnan, C., Pardhi, S. & Karthick, B. 2025. Soil diatoms and their applications as an indicator of environmental changes. *Applied Soil Ecology*, (213):106277.

Zelinka, M., Marvan, P. 1961. Zur Präzisierung der biologischen klassifikation der reinheit fließender gewässer. *Hydrobiolgia*, 57:389-407.

Zheng, Y., Rugenstein, M., Pieper, P., Beobide-Arsuaga, G., & Baehr, J. 2022. El Niño–Southern Oscillation (ENSO) predictability in equilibrated warmer climates. *Earth System Dynamics*, 13(4):1611-1623.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. 2009. Mixed effects models and extensions in ecology with R. New York: Springer.

APPENDICES

Appendix A: Ancillary tables for indices calculated using Inferred Knowledge and Weighted Averaging (WA).

Table A.1: Species optima and tolerances inferred from expert knowledge for nutrients, ionic load and organic load.

Code	Denomination	Genus	(μ) Nutrients	(v) Nutrients	(μ) Ionic load	(v) ionic load	(μ) Organic load	(v) Organic load
AAMB	*Aulacoseira ambigua* (Grunow) Simonsen	AULA	2.0	2.0	2.0	2.0	3.0	2.0
AAMC	*Aulacoseira ambigua* fo. *curvata* Skabicevsky	AULA	1.0	1.0	2.0	2.0	3.0	2.0
ABRY	*Adlafia bryophila* (Petersen) Moser, Lange-Bertalot & Metzeltin	ADLF	4.0	3.0	4.0	3.0	5.0	3.0
ACAF	*Achnanthidium affine* (Grunow) Czarnecki	ACHD	3.0	2.0	3.0	3.0	4.0	3.0
ACHD	*Achnanthidium* Kützing	ACHD	3.1	2.4	3.4	2.9	4.1	2.9
ACHN	*Achnanthes* Bory de St. Vincent	ACHN	4.4	2.6	4.6	2.7	4.7	2.9
ACOF	*Amphora coffeaeformis* (Agardh) Kützing	AMPH	1.0	3.0	1.0	1.0	4.0	2.0
ACOP	*Amphora copulata* (Kützing) Schoeman & Archibald	AMPH	3.0	2.0	3.0	2.0	4.0	3.0
ADCR	*Achnanthidium crassum* (Hustedt) Potapova & Ponader (*Achnanthes crassa*)	ACHD	4.0	2.0	4.0	2.0	5.0	3.0
ADEG	*Achnanthidium exiguum* (Grunow) Czarnecki (*Gogorevia exilis*)	ACHD	3.0	3.0	3.0	3.0	4.0	3.0
ADEU	*Achnanthidium eutrophilum* (Lange-Bertalot) Lange-Bertalot	ACHD	2.0	2.0	3.0	3.0	4.0	3.0
ADLF	*Adlafia* Moser, Lange-Bertalot & Metzeltin	ADLF	4.0	3.0	4.0	3.0	5.0	3.0
ADMA	*Achnanthidium macrocephalum* (Hustedt) Round & Bukhtiyarova	ACHD	5.0	3.0	4.0	3.0	5.0	3.0
ADMI	*Achnanthidium minutissimum* (Kützing) Czarnecki	ACHD	3.0	2.0	3.0	3.0	4.0	3.0
ADRI	*Achnanthidium rivulare* Potapova & Ponader	ACHD	4.0	3.0	4.0	3.0	5.0	3.0
ADSA	*Achnanthidium saprophilum* (Kobayasi & Mayama) Round & Bukhtiyarova	ACHD	1.0	2.0	3.0	3.0	2.0	2.0
ADST	*Achnathidium standeri* (Cholnoky) J.C.Taylor, E.Morales & L.Ector (*Achnanthes standeri*)	ACHD						

AEXG	*Achnanthes exigua* Kützing *(Gogorevia exilis* (Kützing) Kulikovskiy and Kociolek)	ACHN	3.0	2.0	3.0	2.0	4.0	2.0
AFON	*Amphora fontinalis* Hustedt	AMPH	2.0	2.0	3.0	2.0	4.0	3.0
AFOR	*Asterionella formosa* Hassall	ASTE	3.0	2.0	3.0	2.0	4.0	3.0
AINA	*Amphora inariensis* Krammer	AMPH	3.0	3.0	3.0	3.0	4.0	3.0
ADLI	*Achnanthidium linearioides* Lange-Bertalot	ACHN	5.0	3.0	5.0	3.0	5.0	3.0
AMMO	*Amphora montana* Krasske	AMPH	1.0	3.0	3.0	2.0	3.0	3.0
AMPH	*Amphora* Ehrenberg	AMPH	2.1	2.4	2.7	2.0	3.8	2.9
AMPI	*Amphipleura* Kützing	AMPI	3.0	2.0	3.0	2.0	4.0	3.0
AMUZ	*Aulacoseira muzzanensis* (Meister) Krammer	AULA	3.0	2.0	3.0	2.0	4.0	3.0
ANOM	*Anomoeneis* Pfitzer	ANOM	1.0	2.0	1.5	1.5	1.0	2.0
ANOR	*Amphora normanii* Rabenhorst	AMPH	2.0	2.0	3.0	2.0	4.0	3.0
AOBG	Achnanhes oblongella Østrup (*Platessa oblongella* (Østrup) C.E.Wetzel, Lange-Bertalot & Ector)	ACHN	4.0	2.0	5.0	3.0	4.0	3.0
AOVA	*Amphora ovalis* (Kützing) Kützing	AMPH	4.0	3.0	3.0	2.0	4.0	3.0
APED	*Almphora pediculus* (Kützing) Grunow	AMPH	2.0	2.0	3.0	2.0	4.0	3.0
APEL	*Amphipleura pellucida* Kützing	AMPI	3.0	2.0	3.0	2.0	4.0	3.0
ASAF	*Achnanthes subaffinis* Cholnoky	ACHN	5.0	3.0	5.0	3.0	5.0	3.0
ASCO	*Anomoeoneis sphaerophora* fo. *costata* (Kützing)Schmidt	ANOM	1.0	2.0	1.0	1.0	1.0	2.0
ASPH	*Anomoeoneis sphaerophora* (Ehrenberg) Pfitzer	ANOM	1.0	2.0	2.0	2.0	1.0	2.0
ASTE	*Asterionella* A.H. Hassall	ASTE	3.0	2.0	3.0	2.0	4.0	3.0
ASWA	*Achnanthes swazi* Cholnoky	ACHN	5.0	3.0	5.0	3.0	5.0	3.0
AUCS	*Aulacoseira crassipunctata* Krammer	AULA	1.0	3.0	2.0	2.0	3.0	2.0
AUGA	*Aulacoseira granulata* var. *angustissima* (O.Müller) Simonsen	AULA	1.0	3.0	2.0	2.0	2.0	2.0
AUGR	*Aulacoseira granulata* (Ehrenberg) Simonsen	AULA	1.0	3.0	2.0	2.0	3.0	2.0
AULA	*Aulacoseira* Thwaites	AULA	1.6	2.3	2.3	2.0	3.1	2.1

AUSB	*Aulacoseira subborealis* (Nygaard) Denys, Muylaert & Krammer	AULA	2.0	2.0	3.0	2.0	4.0	2.0
AVEN	*Amphora veneta* Kützing	AMPH	1.0	2.0	2.0	2.0	3.0	3.0
BACI	*Bacillaria* Gmelin	BACI	2.0	2.0	1.0	1.0	4.0	3.0
BBRE	*Brachysira brebissonii* Ross	BRAC	5.0	3.0	5.0	3.0	5.0	3.0
BNEO	*Brachysira neoexilis* Lange-Bertalot	BRAC	4.0	3.0	3.0	3.0	5.0	3.0
BPAR	*Bacillaria paradoxa* Gmelin	BACI	2.0	2.0	1.0	1.0	4.0	3.0
BRAC	*Brachysira* Kützing	BRAC	4.6	3.0	4.2	3.0	5.0	3.0
BVIT	*Brachysira vitrea* (Grunow) Ross	BRAC	4.0	3.0	3.0	3.0	5.0	3.0
BWYG	*Brachysira wygaschii* Lange-Bertalot	BRAC	5.0	3.0	5.0	3.0	5.0	3.0
BZEL	*Brachysira zellensis* (Grunow) Round & Mann	BRAC	5.0	3.0	5.0	3.0	5.0	3.0
CACD	*Craticula acidoclinata* Lange-Bertalot & Metzeltin	CRAT	4.0	2.0	4.0	3.0	4.0	3.0
CACM	*Craticula accomodiformis* Lange-Bertalot	CRAT	1.0	2.0	2.0	2.0	3.0	3.0
CALO	*Caloneis* Cleve	CALO	3.2	3.0	3.5	2.8	4.7	3.0
CAMB	*Craticula ambigua* (Ehrenberg) D.G. Mann	CRAT	1.0	3.0	2.0	2.0	2.0	2.0
CAPA	*Capartogramma* Kufferath	CAPA	2.0	2.0	3.0	3.0	4.0	3.0
CAQT	*Caloneis aequatorialis* Hustedt	CALO	3.0	3.0	4.0	2.0	5.0	3.0
CASP	*Cymbella aspera* (Ehrenberg) H. Peragallo	CYMB	4.0	3.0	4.0	3.0	5.0	3.0
CATO	*Cyclotella atomus* Hustedt	CYCL	1.0	3.0	2.0	3.0	2.0	2.0
CBAC	*Caloneis bacillum* (Grunow) Cleve	CALO	2.0	3.0	3.0	3.0	4.0	3.0
CBAM	*Cymbopleura amphicephala* Krammer	CBPL	5.0	3.0	5.0	3.0	5.0	3.0
CBNA	*Cymbopleura naviculiformis* (Auerswald) Krammer	CBPL	5.0	3.0	5.0	3.0	5.0	3.0
CBPL	*Cymbopleura* (Krammer) Krammer	CBPL	5.0	3.0	5.0	3.0	5.0	3.0
CCLY	*Campylodiscus clypeus* Ehrenberg	CPLD	2.0	1.0	1.0	1.0	2.0	2.0
CCRU	*Capartogramma crucicula*(Grunow) Ross	CAPA	2.0	2.0	3.0	3.0	4.0	3.0
CCRY	*Cyclotella cryptica* Reimann & Volcani	CYCL	1.0	3.0	2.0	3.0	2.0	2.0
CCST	*Cyclostephanos* Round	CCST	1.0	1.0	2.0	2.0	2.0	1.0
CCYM	*Cymbella cymbiformis* Agardh	CYMB	3.0	3.0	4.0	3.0	5.0	3.0
CDUB	*Cyclostephanos dubius* (Fricke) Round	CCST	1.0	1.0	2.0	2.0	2.0	1.0

CENG	*Cocconeis engelbrechtii* Cholnoky		2.0	3.0	2.0	2.0	2.0	3.0
CHAL	*Craticula halophila* (Grunow) D.G. Mann	CRAT	2.0	2.0	1.0	1.0	3.0	2.0
CHYA	*Caloneis hyalina* Hustedt	CALO	5.0	3.0	4.0	3.0	5.0	3.0
CINV	*Cyclostephanos invisitatus* (Hohn & Hellerman) Theriot Stoermer & Håkansson	CCST	1.0	1.0	2.0	2.0	2.0	1.0
CKOL	*Cymbella kolbei* Hustedt	CYMB	4.0	3.0	3.0	3.0	5.0	3.0
CKPP	*Cymbella kappii* (Cholnoky) Cholnoky	CYMB	3.0	3.0	3.0	3.0	5.0	3.0
CMED	*Cyclotella meduanae* Germain	CYCL	1.0	1.0	2.0	2.0	4.0	2.0
CMEN	*Cyclotella meneghiniana* Kützing	CYCL	1.0	3.0	2.0	3.0	2.0	2.0
CMLF	*Craticula molestiformis* (Hustedt) Lange-Bertalot	CRAT	1.0	2.0	2.0	1.0	1.0	2.0
CMOL	*Caloneis molaris* (Grunow) Krammer	CALO	2.0	3.0	3.0	3.0	4.0	3.0
CNCI	*Cymbella neocistula* Krammer	CYMB	4.0	3.0	4.0	3.0	5.0	3.0
COCE	*Cyclotella ocellata* Pantocsek	CYCL	2.0	2.0	2.0	2.0	3.0	3.0
COCO	*Cocconeis* Ehrenberg	COCO	2.5	2.5	3.0	2.0	4.0	2.5
CPED	*Cocconeis pediculus* Ehrenberg	COCO	2.0	3.0	3.0	2.0	4.0	3.0
CPLA	*Cocconeis placentula* Ehrenberg	COCO	3.0	2.0	3.0	2.0	4.0	3.0
CPLD	*Campylodiscus* Ehrenberg	CPLD	2.0	1.0	1.0	1.0	2.0	2.0
CPLE	*Cocconeis placentula* Ehrenberg var. *euglypta* (Ehrenberg) Grunow	COCO	2.0	2.0	3.0	2.0	4.0	2.0
CPLI	*Cocconeis placentula* Ehrenberg var. *lineata* (Ehrenberg) Van Heurck	COCO	3.0	3.0	3.0	2.0	4.0	2.0
CRAC	*Craticula accomoda* (Hustedt) Mann	CRAT	1.0	1.0	2.0	2.0	2.0	2.0
CRAT	*Craticula* Grunow	CRAT	2.0	2.3	2.5	2.0	2.9	2.4
CRBU	*Craticula buderi* (Hustedt) Lange-Bertalot	CRAT	1.0	2.0	2.0	1.0	2.0	2.0
CRCU	*Craticula cuspidata* (Kützing) D.G. Mann	CRAT	1.0	3.0	2.0	2.0	2.0	2.0
CSAP	*Cymatopleura solea* var. *apiculata* (W. Smith) Ralfs	CYMA	1.0	2.0	2.0	2.0	2.0	2.0
CSBM	*Craticula submolesta* (Hustedt) Lange-Bertalot	CRAT	4.0	3.0	4.0	3.0	5.0	3.0
CSHU	*Caloneis schumanniana* (Grunow) Cleve	CALO	3.0	3.0	3.0	3.0	5.0	3.0
CSIL	*Caloneis silicula* (Ehrenberg) Cleve	CALO	4.0	3.0	4.0	3.0	5.0	3.0
CSLP	*Cymbella subleptoceros* Krammer	CYMB	4.0	3.0	4.0	3.0	5.0	3.0

CSMO	*Cymbella simonsenii* Krammer	CYMB	4.0	3.0	4.0	3.0	5.0	3.0
CSOL	*Cymatopleura solea* (Brébisson) W. Smith	CYMA	1.0	2.0	2.0	2.0	2.0	2.0
CSTT	*Discostella stelligera* (Cleve and Grunow) Houk & Klee	CYCL	2.0	2.0	3.0	2.0	3.0	1.0
CTGL	*Cymbella turgidula* Grunow	CYMB	3.0	3.0	3.0	3.0	5.0	3.0
CTNP	Ctenophora (A. Grunow) D.M. Williams & Round	CTNP	1.0	2.0	2.0	1.0	2.0	2.0
CTPU	*Ctenophora pulchella* (Ralfs) Williams & Round	CTNP	1.0	2.0	2.0	1.0	2.0	2.0
CTUM	*Cymbella tumida* (Brébisson) Van Heurck	CYMB	3.0	3.0	3.0	3.0	5.0	3.0
CVIX	*Craticula vixnegligenda* Lange-Bertalot	CRAT	4.0	3.0	4.0	3.0	5.0	3.0
CWOL	*Cyclotella woltereckii* Hustedt	CYCL	2.0	2.0	3.0	2.0	3.0	1.0
CYCL	*Cyclotella* Kützing	CYCL	1.4	2.3	2.3	2.4	2.7	1.9
CYMA	*Cymatopleura* W. Smith	CYMA	1.0	2.0	2.0	2.0	2.0	2.0
CYMB	*Cymbella* C.Agardh	CYMB	3.6	3.0	3.6	3.0	5.0	3.0
DCOF	*Diadesmis confervacea* Kützing	DIAM	1.0	1.0	2.0	2.0	4.0	2.0
DENT	*Denticula* Kützing	DENT	2.0	2.7	2.0	1.7	4.3	3.0
DIAM	*Diadesmis* Kützing	DIAM	1.0	1.0	2.0	2.0	4.0	2.0
DIAT	*Diatoma* Bory de St. Vincent	DIAT	1.0	2.0	3.0	2.0	4.0	3.0
DIPL	*Diploneis* Ehrenberg	DIPL	2.0	3.0	3.0	2.0	4.0	3.0
DISC	*Discostella* Houk & Klee	DISC	2.0	2.0	3.0	2.0	3.0	1.0
DKUE	*Denticula kuetzingii* Grunow	DENT	2.0	3.0	2.0	2.0	5.0	3.0
DPST	*Discostella pseudostelligera* (Hustedt) Houk & Klee	DISC	2.0	2.0	3.0	2.0	3.0	1.0
DSMI	*Diploneis smithii* (Brébisson) Cleve	DIPL	2.0	2.0	1.0	2.0	3.0	2.0
DSTE	*Discostella stelligera* (Cleve & Grunow) Houk & Klee	DISC	2.0	2.0	3.0	2.0	3.0	1.0
DSUB	*Denticula subtilis* Grunow	DENT	2.0	3.0	2.0	2.0	4.0	3.0
DSUN	*Denticula sundayensis* Archibald	DENT	2.0	2.0	2.0	1.0	4.0	3.0
DVUL	*Diatoma vulgaris* Bory	DIAT	1.0	2.0	3.0	2.0	4.0	3.0
DWOL	*Discostella woltereckii* (Hustedt) Houk & Klee	DISC	2.0	2.0	3.0	2.0	3.0	1.0
EADN	*Epithemia adnata* (Kützing) Brébisson	EPIT	2.0	3.0	3.0	3.0	4.0	3.0
EBIL	*Eunotia bilunaris* (Ehrenberg) Mills	EUNO	4.0	3.0	4.0	3.0	5.0	3.0
ECAE	*Encyonema caespitosum* Kützing	ENCY	4.0	3.0	4.0	3.0	5.0	3.0

ECBU	*Encyonopsis buedelii* Krammer	ENCP	5.0	3.0	5.0	3.0	5.0	3.0
ECES	*Encyonopsis cesatii* (Rabenhorst) Krammer	ENCP	5.0	3.0	5.0	3.0	5.0	3.0
ECFA	*Encyonopsis falaisensis* (Grunow) Krammer	ENCP	5.0	3.0	5.0	3.0	5.0	3.0
ECKR	*Encyonopsis krammeri* Reichardt	ENCP	4.0	3.0	4.0	3.0	5.0	3.0
ECPM	*Encyonopsis minuta* Krammer & Reichardt	ENCP	3.0	3.0	3.0	3.0	4.0	3.0
EEXI	*Eunotia exigua* (Brébisson) Rabenhorst	EUNO	5.0	3.0	5.0	3.0	5.0	3.0
EFLE	*Eunotia flexuosa* (Brébisson) Kützing	EUNO	5.0	3.0	5.0	3.0	5.0	3.0
EFOR	*Eunotia formica* Ehrenberg	EUNO	4.0	3.0	4.0	3.0	5.0	2.0
EINC	*Eunotia incisa* Gregory	EUNO	4.0	3.0	4.0	3.0	5.0	3.0
EMIN	*Eunotia minor* (Kützing) Grunow	EUNO	5.0	3.0	4.0	3.0	5.0	3.0
ENCM	*Encyonopsis microcephala* (Grunow) Krammer	ENCP	4.0	3.0	4.0	2.0	5.0	3.0
ENCP	Encyonopsis Krammer	ENCP	4.0	3.0	4.2	2.9	4.8	3.0
ENCY	*Encyonema* Kützing	ENCY	4.2	2.8	4.0	2.8	5.0	3.0
ENLS	*Encyonopsis leei* Krammer	ENCP	2.0	3.0	3.0	3.0	4.0	3.0
ENME	*Encyonema mesianum* (Cholnoky) D.G. Mann	ENCY	4.0	3.0	4.0	3.0	5.0	3.0
ENMI	*Encyonema minutum* (Hilse) D.G. Mann	ENCY	4.0	3.0	4.0	2.0	5.0	3.0
ENNG	*Encyonema neogracile* Krammer	ENCY	5.0	2.0	4.0	3.0	5.0	3.0
ENVE	*Encyonema ventricosum* (Agardh) Grunow	ENCY	4.0	3.0	4.0	3.0	5.0	3.0
EOAR	*Eolimna archibaldii* Taylor & Lange-Bertalot (*Sellaphora archibaldii* (Taylor & Lange-Bertalot) Ács, C.E. Wetzel & Ector)	EOLI	1.0	1.0	2.0	1.0	2.0	2.0
EOLI	*Eolimna* Lange-Bertalot & Schiller	EOLI	1.0	1.0	2.0	1.7	2.0	1.7
EOMI	*Eolimna minima* (Grunow) Lange-Bertalot (*Sellaphora atomoides*)	EOLI	1.0	1.0	2.0	2.0	2.0	2.0
EPIT	*Epithemia* Kützing	EPIT	2.0	3.0	3.5	3.0	4.0	3.0
EPUN	*Eunotia pectinalis* var. *undulata* (Ralfs) Rabenhorst	EUNO	4.0	3.0	4.0	3.0	5.0	3.0
ERAY	*Encyonopsis raytonensis* (Cholnoky) Krammer	ENCP	4.0	3.0	5.0	3.0	5.0	3.0
ERHO	*Eunotia rhomboidea* Hustedt	EUNO	5.0	3.0	5.0	3.0	5.0	3.0
ESBM	*Eolimna subminuscula* (Manguin) Moser, Lange-	EOLI	1.0	1.0	2.0	2.0	2.0	1.0

	Bertalot & Metzeltin (*Craticula subminiscula*)							
ESLE	*Encyonema silesiacum* (Bleisch) D.G. Mann	ENCY	4.0	3.0	4.0	3.0	5.0	3.0
ESOR	*Epithemia sorex* Kützing	EPIT	2.0	3.0	4.0	3.0	4.0	3.0
ESUM	*Encyonopsis subminuta* Krammer & Reichardt	ENCP	4.0	3.0	4.0	3.0	5.0	3.0
ETEN	*Eunotia tenella* (Grunow)Hustedt	EUNO	5.0	3.0	5.0	3.0	5.0	3.0
EUNO	*Eunotia* Ehrenberg	EUNO	4.6	3.0	4.4	3.0	5.0	2.9
FALL	*Fallacia* Stickle & D.G. Mann	FALL	1.7	1.8	2.2	2.3	3.2	2.5
FBCP	*Fragilaria biceps* (Kützing) Lange-Bertalot (*Ulnaria biceps*)	FRAG	2.0	2.0	3.0	2.0	4.0	3.0
FCAP	*Fragilaria capucina* Desmazieres	FRAG	3.0	2.0	3.0	2.0	4.0	3.0
FCRO	*Fragilaria crotonensis* Kitton	FRAG	3.0	2.0	3.0	2.0	4.0	3.0
FCRS	*Frustulia crassinervia* (Brébisson) Lange-Bertalot & Krammer	FRUS	5.0	3.0	4.0	3.0	5.0	3.0
FCRU	*Fragilaria capucina* var. *rumpens* (Kützing) Lange-Bertalot (*Fragilaria rumpens*)	FRAG	3.0	2.0	3.0	2.0	4.0	3.0
FCVA	*Fragilaria capucina* Desmazieres var. *vaucheriae* (Kützing) Lange-Bertalot (*Fragilaria vaucheriae*)	FRAG	3.0	2.0	3.0	2.0	4.0	3.0
FINS	*Fallacia insociabilis* (Krasske) D.G. Mann	FALL	2.0	2.0	3.0	3.0	4.0	3.0
FITU	*Fistulifera* Lange-Bertalot	FITU	1.0	1.0	2.0	1.0	1.0	1.0
FMGL	*Frustulia magaliesmontana* Cholnoky	FRUS	5.0	3.0	5.0	3.0	5.0	3.0
FMOC	*Fallacia monoculata* (Hustedt) D.G. Mann	FALL	2.0	2.0	2.0	3.0	4.0	3.0
FNAN	*Fragilaria nanana* Lange-Bertalot	FRAG	4.0	3.0	3.0	2.0	4.0	3.0
FPMG	*Frustulia pseudomagaliesmontana* Camburn & Charles	FRUS	5.0	3.0	5.0	3.0	5.0	3.0
FPSC	*Fragilaria parasitica* var. *subconstricta* Grunow	FRAG	2.0	3.0	3.0	2.0	4.0	3.0
FPYG	*Fallacia pygmaea* (Kützing) Stickle & D.G. Mann	FALL	1.0	1.0	1.0	1.0	1.0	1.0
FRAG	*Fragilaria* Lyngbye	FRAG	2.7	2.3	2.8	2.0	3.9	2.7
FROS	*Frustulia rostrata* Hustedt	FRUS	4.0	3.0	4.0	3.0	5.0	3.0
FRUM	*Fragilaria rumpens* (Kützing) G.W.F.Carlson	FRAG	4.0	3.0	3.0	2.0	4.0	3.0
FRUS	*Frustulia* Rabenhorst	FRUS	4.6	3.0	4.4	3.0	4.9	3.0

FSAP	*Fistulifera saprophila* (Lange-Bertalot & Bonik) Lange-Bertalot	FITU	1.0	1.0	2.0	1.0	1.0	1.0
FSAX	*Frustulia saxonica* Rabenhorst	FRUS	4.0	3.0	4.0	3.0	5.0	3.0
FSBH	*Fallacia subhamulata* (Grunow) D.G. Mann	FALL	3.0	2.0	3.0	3.0	4.0	3.0
FTEN	*Fragilaria tenera* (W. Smith) Lange-Bertalot	FRAG	2.0	2.0	3.0	2.0	4.0	2.0
FTNR	*Fallacia tenera* (Hustedt) Mann in Round	FALL	1.0	2.0	2.0	2.0	3.0	2.0
FTUG	*Frustulia tugelae* Cholnoky	FRUS	5.0	3.0	5.0	3.0	5.0	3.0
FUAC	*Fragilaria ulna* var. *acus* (Kützing) Lange-Bertalot (*Ulnaria acus*)	FRAG	2.0	2.0	2.0	2.0	4.0	2.0
FUMP	*Fallacia umpatica* (Cholnoky) D.G. Mann	FALL	1.0	2.0	2.0	2.0	3.0	3.0
FVAU	*Fragilaria vaucheriae* (Kützing) Petersen	FRAG	2.0	2.0	2.0	2.0	3.0	2.0
FVUL	*Frustulia vulgaris* (Thwaites) De Toni	FRUS	4.0	3.0	3.0	3.0	4.0	3.0
FWEI	*Frustulia weinholdii* Hustedt	FRUS	5.0	3.0	5.0	3.0	5.0	3.0
GACU	*Gomphonema acuminatum* Ehrenberg	GOMP	3.0	3.0	3.0	3.0	5.0	2.0
GAFF	*Gomphonema affine* Kützing	GOMP	4.0	3.0	3.0	3.0	4.0	3.0
GANG	*Gomphonema angustatum* (Kützing) Rabenhorst	GOMP	3.0	3.0	3.0	3.0	5.0	3.0
GCAP	*Gomphonema capitatum* Ehrenberg	GOMP	2.0	3.0	3.0	3.0	5.0	3.0
GCLA	*Gomphonema clavatum* Ehrenberg	GOMP	2.0	3.0	3.0	3.0	4.0	3.0
GDEC	*Geissleria decussis* (Østrup) Lange-Bertalot & Metzeltin	GEIS	2.0	2.0	2.0	3.0	4.0	2.0
GEIS	*Geissleria* Lange-Bertalot & Metzeltin	GEIS	2.0	2.0	2.0	3.0	4.0	2.0
GEXL	*Gomphonema exilissimum* (Grunow) Lange-Bertalot & Reichardt	GOMP	3.0	3.0	3.0	3.0	5.0	3.0
GGRA	*Gomphonema gracile* Ehrenberg	GOMP	3.0	3.0	3.0	3.0	5.0	3.0
GINS	*Gomphonema insigne* Gregory	GOMP	2.0	3.0	3.0	3.0	4.0	3.0
GITA	*Gomphonema italicum* Kützing	GOMP	1.0	3.0	3.0	3.0	4.0	3.0
GLTC	*Gomphonema laticollum* Reichardt	GOMP	1.0	3.0	3.0	3.0	4.0	3.0
GMIN	Gomphonema minutum (Agardh) Agardh	GOMP	2.0	3.0	3.0	3.0	5.0	3.0

GOMP	*Gomphonema* Ehrenberg	GOMP	2.0	2.8	3.1	2.9	4.3	2.9
GPAR	*Gomphonema parvulum* (Kützing) Kützing	GOMP	1.0	1.0	3.0	2.0	3.0	2.0
GPAS	*Gomphonema parvulum* var. *parvulum* f. *saprophilum* Lange-Bert.&Reichardt	GOMP	1.0	1.0	3.0	2.0	2.0	2.0
GPLA	*Gomphonema parvulum* var. *lagenula* (Kützing) Frenguelli	GOMP	1.0	3.0	3.0	3.0	4.0	3.0
GPPA	*Gomphonema parvulum* var. *parvulius* Lange-Bertalot & Reichardt	GOMP	2.0	3.0	4.0	3.0	5.0	3.0
GPRI	*Gomphonema pumilum* var. *rigidum* Reichardt & Lange-Bertalot	GOMP	1.0	3.0	3.0	3.0	4.0	3.0
GPSA	*Gomphonema pseudoaugur* Lange-Bertalot	GOMP	1.0	2.0	3.0	3.0	3.0	3.0
GPUM	*Gomphonema pumilum* (Grunow) Reichardt & Lange-Bertalot	GOMP	2.0	3.0	3.0	3.0	5.0	3.0
GRAU	*Gyrosigma rautenbachiae* Cholnoky	GYRO	3.0	2.0	3.0	3.0	4.0	3.0
GSCA	*Gyrosigma scalproides* (Rabenhorst)Cleve	GYRO	1.0	2.0	2.0	2.0	4.0	2.0
GTRU	*Gomphonema truncatum* Ehrenberg	GOMP	2.0	3.0	3.0	3.0	5.0	3.0
GVNU	*Gomphonema venusta* Passy, Kociolek & Lowe	GOMP	3.0	3.0	3.0	3.0	5.0	3.0
GYAC	*Gyrosigma acuminatum* (Kützing) Rabenhorst	GYRO	2.0	3.0	3.0	3.0	4.0	3.0
GYAT	*Gyrosigma attenuatum* (Kützing) Rabenhorst	GYRO	2.0	3.0	3.0	3.0	4.0	3.0
GYRO	*Gyrosigma* Hassall	GYRO	2.0	2.5	2.8	2.8	4.0	2.8
HACO	*Halamphora coffeaeformis* (Agardh) Levkov	HALA	1.0	3.0	1.0	1.0	4.0	2.0
HALA	*Halamphora* (Cleve) Levkov	HALA	1.5	2.5	2.0	1.5	4.0	2.5
HANT	*Hantzschia* Grunow	HANT	2.0	2.0	1.0	1.0	4.0	3.0
HAVT	*Hippodonta avittata* (Cholnoky) Lange-Bertalot, Metzeltin & Witkowski	HIPO	1.0	2.0	1.0	2.0	3.0	2.0
HCAP	*Hippodonta capitata* (Ehrenberg) Lange-Bertalot, Metzeltin & Witkowski	HIPO	1.0	2.0	2.0	2.0	3.0	2.0
HDIS	*Hantzschia distinctepunctata* Hustedt	HANT	2.0	2.0	1.0	1.0	4.0	3.0

HIPO	*Hippodonta* Lange-Bertalot, Metzeltin & Witkowski	HIPO	1.0	2.0	1.5	2.0	3.0	2.0
HNOR	*Halamphora normanii* (Rabenhorst) Levkov	HALA	2.0	2.0	3.0	2.0	4.0	3.0
KOBL	*Kobayasiella* Lange-Bertalot	KOBL	4.0	3.0	5.0	3.0	5.0	3.0
KOSU	*Kobayasiella subtilissima* (Cleve) Lange-Bertalot	KOBL	4.0	3.0	5.0	3.0	5.0	3.0
LEMN	*Lemnicola* Round & Basson	LEMN	2.0	3.0	3.0	3.0	4.0	3.0
LHUN	*Lemnicola hungarica* (Grunow) Round & Basson	LEMN	2.0	3.0	3.0	3.0	4.0	3.0
MAAT	*Mayamaea atomus* (Kützing) Lange-Bertalot	MAYA	1.0	1.0	2.0	2.0	1.0	1.0
MAPE	*Mayamaea atomus* var. *permitis* (Hustedt) Lange-Bertalot	MAYA	1.0	1.0	2.0	2.0	1.0	1.0
MAST	*Mastogloia* Thwaites	MAST	2.0	3.0	1.5	1.0	4.0	3.0
MAYA	*Mayamaea* Lange-Bertalot	MAYA	1.0	1.0	2.0	2.0	1.0	1.0
MDAN	*Mastogloia dansei* (Thwaites) W.M.Smith	MAST	2.0	2.0	1.0	2.0	4.0	3.0
MELL	*Mastogloia elliptica* (C.A. Agardh) Cleve	MAST	2.0	3.0	2.0	1.0	4.0	3.0
MELO	*Melosira* Agardh	MELO	2.0	3.0	3.0	2.0	4.0	2.0
MSMI	*Mastogloia smithii* Thwaites	MAST	2.0	3.0	1.0	1.0	4.0	3.0
MVAR	*Melosira varians* Agardh	MELO	2.0	3.0	3.0	2.0	4.0	2.0
NAAN	*Navicula angusta* Grunow	NAVI	4.0	3.0	3.0	3.0	5.0	3.0
NACD	*Nitzschia acidoclinata* Lange-Bertalot	NITZ	4.0	3.0	4.0	3.0	4.0	3.0
NACI	*Nitzschia acicularis* (Kützing) W.M.Smith	NITZ	1.0	1.0	2.0	2.0	1.0	2.0
NAGN	*Nitzschia agnita* Hustedt	NITZ	1.0	2.0	3.0	3.0	3.0	2.0
NAGW	*Nitzschia agnewii* Cholnoky	NITZ	1.0	2.0	3.0	1.0	3.0	1.0
NAMA	*Navicula arvensis* Hustedt var. *maior* Lange-Bertalot	NAVI	1.0	2.0	2.0	2.0	2.0	2.0
NAMC	*Nitzschia amplectens* Hustedt	NITZ	4.0	1.0	1.0	3.0	4.0	1.0
NAMP	*Nitzschia amphibia* Grunow	NITZ	1.0	1.0	3.0	1.0	2.0	1.0
NANT	*Navicula antonii* Lange-Bertalot	NAVI	1.0	2.0	2.0	2.0	4.0	2.0
NAUR	*Nitzschia aurariae* Cholnoky	NITZ	1.0	1.0	2.0	2.0	2.0	2.0
NAVG	*Navigiolum* Lange-Bertalot, Cavacini. Tagliaventi & Alfinito	NAVG	1.0	1.0	2.0	1.0	2.0	2.0
NAVI	*Navicula* Bory de St. Vincent	NAVI	1.8	2.3	2.5	2.3	3.8	2.5

NBCL	*Nitzschia bacillum* Hustedt	NITZ	4.0	3.0	2.0	2.0	4.0	2.0
NCIN	*Navicula cincta* (Ehrenberg) Ralfs	NAVI	2.0	3.0	2.0	2.0	4.0	2.0
NCLA	*Nitzschia clausii* Hantzsch	NITZ	1.0	2.0	3.0	2.0	3.0	2.0
NCOM	*Nitzschia communis* Rabenhorst	NITZ	1.0	1.0	2.0	2.0	1.0	1.0
NCPL	*Nitzschia capitellata* Hustedt	NITZ	1.0	1.0	3.0	1.0	2.0	1.0
NCPR	*Navicula capitatoradiata* Germain	NAVI	1.0	3.0	3.0	2.0	4.0	2.0
NCPU	*Navicymbula pusilla* Krammer	NVCB	3.0	3.0	2.0	3.0	4.0	3.0
NCRY	*Navicula cryptocephala* Kützing	NAVI	1.0	2.0	2.0	2.0	3.0	2.0
NCTE	*Navicula cryptotenella* Lange-Bertalot	NAVI	1.0	2.0	2.0	2.0	3.0	2.0
NCTO	*Navicula cryptotenelloides* Lange-Bertalot	NAVI	2.0	2.0	2.0	2.0	4.0	3.0
NDES	*Nitzschia desertorum* Hustedt	NITZ	1.0	1.0	3.0	1.0	2.0	1.0
NDIS	*Nitzschia dissipatea* (Kützing) Grunow	NIDI	2.0	3.0	3.0	2.0	5.0	3.0
NDME	*Nitzschia dissipata* var. *media* (Hantzsch) Grunow	NIDI	2.0	3.0	3.0	2.0	5.0	3.0
NDOI	*Navicula dutoitana* (Cholnoky)	NAVI	5.0	3.0	5.0	3.0	5.0	3.0
NDRA	*Nitzschia draveillensis* Coste & Ricard	NITZ	1.0	1.0	2.0	2.0	2.0	2.0
NEAF	*Neidium affine* (Ehrenberg) Pfitzer	NEID	3.0	3.0	4.0	3.0	5.0	3.0
NEID	*Neidium* Pfitzer	NEID	3.5	3.0	4.0	3.0	4.5	3.0
NELE	*Nitzschia elegantula* Grunow	NITZ	1.0	2.0	2.0	1.0	2.0	2.0
NEPR	*Neidium productum* (W.M.Smith) Cleve	NEID	4.0	3.0	4.0	3.0	4.0	3.0
NERI	*Navicula erifuga* Lange-Bertalot	NAVI	1.0	1.0	2.0	2.0	3.0	1.0
NETO	*Nitzschia etoshensis* Cholnoky	NITZ	2.0	2.0	1.0	2.0	3.0	1.0
NFIL	*Nitzschia filiformis* (W.M.Smith) Van Heurck	NITZ	1.0	1.0	3.0	2.0	2.0	1.0
NFON	*Nitzschia fonticola* (Grunow) Grunow	NITZ	2.0	2.0	3.0	2.0	3.0	2.0
NGAD	*Navigiolum adamantiforme* (Archibald) Taylor & Lange-Bertalot	NAVG	1.0	1.0	2.0	1.0	2.0	2.0
NGER	*Navicula germainii* Wallace	NAVI	1.0	2.0	2.0	3.0	4.0	3.0
NGRE	*Navicula gregaria* Donkin	NAVI	1.0	2.0	2.0	1.0	3.0	2.0
NHAN	*Nitzschia hantzschiana* Rabenhorst	NITZ	4.0	3.0	4.0	3.0	4.0	3.0

NHEU	*Nitzschia heufleriana* Grunow	NITZ	2.0	2.0	3.0	2.0	4.0	2.0
NHMD	*Navicula heimansioides* Lange-Bertalot	NAVI	4.0	3.0	3.0	3.0	5.0	3.0
NIAR	*Nitzschia archibaldii* Lange-Bertalot	NITZ	1.0	2.0	3.0	2.0	2.0	2.0
NIDI	Nitzschiae dissipatae (Section)	NIDI	2.0	3.0	3.0	2.0	5.0	3.0
NIFR	*Nitzschia frustulum* (Kützing) Grunow	NITZ	1.0	1.0	2.0	1.0	2.0	1.0
NIGR	*Nitzschia gracilis* Hantzsch	NITZ	1.0	2.0	2.0	2.0	3.0	3.0
NINT	*Nitzschia intermedia* Hantzsch	NITZ	1.0	2.0	3.0	2.0	3.0	2.0
NIPR	*Nitzschia pura* Hustedt	NITZ	1.0	2.0	3.0	2.0	4.0	2.0
NIPU	*Nitzschia pusilla* (Kützing) Grunow	NITZ	1.0	2.0	3.0	2.0	3.0	3.0
NIRM	*Nitzschia irremissa* Cholnoky	NITZ	1.0	2.0	3.0	2.0	3.0	2.0
NITZ	*Nitzschia* Hassall	NITZ	1.4	1.8	2.5	1.7	2.7	1.8
NKUZ	*Nitzschia kurzii* Rabenhorst	NITZ	1.0	2.0	1.0	1.0	2.0	1.0
NLBT	*Nitzschia liebetruthii* Rabenhorst	NITZ	1.0	1.0	2.0	1.0	2.0	1.0
NLGC	*Navicula longicephala* Hustedt	NAVI	2.0	3.0	3.0	3.0	5.0	3.0
NLIB	*Navicula libonensis* Schoeman	NAVI	1.0	2.0	3.0	2.0	4.0	2.0
NLIN	*Nitzschia linearis* (Agardh) W.M. Smith	NITZ	1.0	2.0	3.0	2.0	3.0	2.0
NLSU	*Nitzschia linearis* var. *subtilis* (Grunow) Hustedt	NITZ	1.0	2.0	3.0	2.0	3.0	2.0
NLTL	*Nitzschia lancettula* O.Müller	NITZ	3.0	2.0	3.0	2.0	4.0	3.0
NLTT	*Nitzschia littorea* Grunow	NITZ	1.0	2.0	1.0	1.0	2.0	1.0
NMCA	*Navicula microcari* Lange-Bertalot	NAVI	2.0	3.0	2.0	2.0	4.0	3.0
NMCB	*Navicula microrhombus* (Cholnoky) Schoeman & Archibald	NAVI	1.0	2.0	2.0	2.0	2.0	2.0
NMIC	*Nitzschia microcephala* Grunow	NITZ	1.0	2.0	2.0	1.0	2.0	2.0
NNAN	*Nitzschia nana* Grunow	NITZ	2.0	2.0	3.0	2.0	4.0	2.0
NNOT	*Navicula notha* Wallace	NAVI	4.0	3.0	4.0	2.0	5.0	3.0
NPAE	*Nitzschia paleacea* (Grunow) Grunow	NITZ	1.0	2.0	3.0	2.0	4.0	2.0
NPAL	*Nitzschia palea* (Kützing) W. Smith	NITZ	1.0	1.0	2.0	1.0	1.0	1.0
NPML	*Nitzschia pumila* Hustedt	NITZ	1.0	2.0	3.0	2.0	3.0	2.0
NPRP	*Nitzschia perspicua* Cholnoky	NITZ	2.0	2.0	1.0	2.0	3.0	1.0
NRAD	*Navicula radiosa* Kützing	NAVI	1.0	3.0	2.0	1.0	4.0	3.0

NRAN	*Navicula ranomafanensis* (Manguin) Metzeltin & Lange-Bertalot	NAVI	3.0	3.0	3.0	3.0	4.0	3.0
NRCH	*Navicula reichardtiana* Lange-Bertalot	NAVI	2.0	2.0	3.0	2.0	4.0	3.0
NRCS	*Navicula recens* (Lange-Bertalot) Lange-Bertalot	NAVI	1.0	2.0	2.0	2.0	3.0	2.0
NREC	*Nitzschia recta* Hantzsch	NITZ	2.0	2.0	3.0	2.0	4.0	2.0
NREV	*Nitzschia reversa* W. Smith	NITZ	1.0	2.0	2.0	1.0	2.0	2.0
NRHY	*Navicula rhynchocephala* Kützing	NAVI	2.0	3.0	2.0	3.0	4.0	3.0
NRIE	*Navicula riediana* Lange-Bertalot & Rumrich	NAVI	1.0	2.0	2.0	2.0	3.0	2.0
NROS	*Navicula rostellata* Kützing	NAVI	1.0	2.0	2.0	2.0	4.0	2.0
NSBL	*Nitzschia sublinearis* Hustedt	NITZ	1.0	2.0	3.0	2.0	3.0	2.0
NSDE	*Nitzschia sinuata* var. *delognei* (Grunow) Lange-Bertalot	NITZ	2.0	2.0	3.0	2.0	3.0	2.0
NSHR	*Navicula schroeteri* Meister	NAVI	1.0	2.0	2.0	2.0	4.0	2.0
NSIG	*Nitzschia sigma* (Kützing) W.M. Smith	NITZ	1.0	2.0	3.0	2.0	2.0	2.0
NSIT	*Nitzschia sinuata* var. *tabellaria* Grunow	NITZ	2.0	2.0	3.0	2.0	3.0	2.0
NSRH	*Navicula subrhynchocephala* Hustedt	NAVI	2.0	2.0	2.0	3.0	4.0	3.0
NSSY	*Navicula schroeteri* var. *symmetrica* (Patrick) Lange-Bertalot	NAVI	1.0	2.0	2.0	2.0	4.0	2.0
NSYM	*Navicula symmetrica* Patrick	NAVI	1.0	2.0	2.0	2.0	4.0	2.0
NTEN	*Navicula tenelloides* Hustedt	NAVI	1.0	3.0	3.0	3.0	4.0	3.0
NTPT	*Navicula tripunctata* (O.F.Müller) Bory	NAVI	2.0	2.0	3.0	2.0	3.0	3.0
NTRV	*Navicula trivialis* Lange-Bertalot	NAVI	2.0	2.0	3.0	2.0	4.0	3.0
NUMB	*Nitzschia umbonata* (Ehrenberg) Lange-Bertalot	NITZ	1.0	1.0	2.0	1.0	1.0	1.0
NVCB	*Navicymbula* Krammer	NVCB	3.0	3.0	2.0	3.0	4.0	3.0
NVDA	*Navicula vandamii* Schoeman & Archibald	NAVI	1.0	2.0	2.0	2.0	3.0	3.0
NVEN	*Navicula veneta* Kützing	NAVI	1.0	1.0	2.0	2.0	2.0	2.0
NVIR	*Navicula viridula* (Kützing) Ehrenberg	NAVI	3.0	2.0	3.0	3.0	4.0	3.0
NVLC	*Nitzschia valdecostata* Lange-Bertalot & Simonsen	NITZ	1.0	2.0	3.0	1.0	3.0	2.0
NZAN	*Navicula zanoni* Hustedt	NAVI	2.0	3.0	3.0	3.0	5.0	3.0

NZCL	*Nitzschia closterium* (Ehrenberg) W. Smith	NITZ	1.0	2.0	1.0	1.0	2.0	1.0
NZRA	*Nitzschia radicula* Hustedt	NITZ	1.0	2.0	3.0	1.0	3.0	2.0
NZSU	*Nitzschia supralitorea* Lange-Bertalot	NITZ	1.0	1.0	3.0	1.0	2.0	2.0
PACR	*Pinnularia acrospheria* W. Smith	PINU	4.0	3.0	4.0	3.0	5.0	3.0
PCLT	*Placoneis clementis* (Grunow) Cox	PLAC	1.0	3.0	1.0	2.0	4.0	3.0
PDIC	*Placoneis dicephala* (W. Smith) Mereschkowsky	PLAC	1.0	3.0	2.0	2.0	4.0	2.0
PDIV	*Pinnularia divergens* W.M.Smith	PINU	5.0	3.0	5.0	3.0	5.0	3.0
PDUN	*Pinnularia divergens* var. *undulata* (Peragello & Heribaud) Hustedt	PINU	5.0	3.0	5.0	3.0	5.0	3.0
PELG	*Placoneis elginensis* (Gregory) Cox	PLAC	2.0	3.0	3.0	3.0	4.0	3.0
PELO	*Pleurosigma elongatum* W.Smith	PLSG	1.0	2.0	2.0	2.0	3.0	2.0
PGIB	*Pinnularia gibba* Ehrenberg	PINU	4.0	3.0	4.0	3.0	5.0	3.0
PINU	*Pinnularia* Ehrenberg	PINU	4.3	2.9	4.4	3.0	4.8	3.0
PJOC	*Pinnularia joculata* (Manguin) Krammer	PINU	5.0	3.0	4.0	3.0	5.0	3.0
PLAC	*Placoneis* Mereschkowsky	PLAC	1.3	3.0	2.0	2.0	3.8	2.5
PLEN	*Planothidium engelbrechtii* (Cholnoky) Round & Bukhtiyarova	PLTD	3.0	3.0	2.0	2.0	4.0	3.0
PLFR	*Planothidium frequentissimum* (Lange-Bertalot) Lange-Bertalot	PLTD	2.0	2.0	3.0	2.0	3.0	2.0
PLHU	*Platessa hustedtii* (Krasske) Lange-Bertalot	PTSA	4.0	2.0	5.0	3.0	4.0	3.0
PLSG	*Pleurosigma* W. Smith	PLSG	1.0	2.0	2.0	1.5	3.0	2.0
PLTD	*Planothidium* Round & Bukhtiyarova	PLTD	2.8	2.3	2.8	2.0	3.8	2.3
PMRO	*Pinnularia microstauron* (Ehrenberg) Cleve var. *rostrata* Krammer	PINU	4.0	3.0	5.0	3.0	5.0	3.0
POBG	*Psammothidium oblongellum* (Oestrup) Van de Vijver (*Platessa oblongella*)	PSMT	3.0	2.0	3.0	3.0	4.0	3.0
PPLC	*Placoneis placentula* (Ehrenberg) Heinzerling	PLAC	1.0	3.0	2.0	1.0	3.0	2.0
PRST	*Planothidium rostratum* (Østrup) Lange-Bertalot	PLTD	3.0	2.0	3.0	2.0	4.0	2.0
PSAL	*Pleurosigma salinarum* (Grunow) Cleve & Grunow	PLSG	1.0	2.0	2.0	1.0	3.0	2.0
PSBR	*Pseudostaurosira brevistriata* (Grunow) D.M. Williams & Round	PSST	2.0	2.0	1.0	2.0	3.0	2.0

PSBV	*Pinnularia subbrevistriata* Krammer	PINU	2.0	2.0	3.0	3.0	3.0	3.0
PSCA	*Pinnularia subcapitata* Gregory	PINU	4.0	3.0	4.0	3.0	5.0	3.0
PSMT	*Psammothidium* Bukhtiyarova & Round	PSMT	3.0	2.0	3.0	3.0	4.0	3.0
PSST	*Pseudostaurosira* (Grunow) D.M. Williams & Round	PSST	2.0	2.0	1.0	2.0	3.0	2.0
PTLA	*Planothidium lanceolatum* (Brébisson) Lange-Bertalot	PLTD	3.0	2.0	3.0	2.0	4.0	2.0
PTSA	*Platessa* Lange Bertalot	PTSA	4.0	2.0	5.0	3.0	4.0	3.0
PVIF	*Pinnularia viridiformis* Krammer	PINU	5.0	3.0	5.0	3.0	5.0	3.0
PVIR	*Pinnularia viridis* (Nitzsch) Ehrenberg	PINU	5.0	3.0	5.0	3.0	5.0	3.0
RABB	*Rhoicosphenia abbreviata* (C. Agardh) Lange-Bertalot	RHOI	1.0	3.0	3.0	3.0	4.0	3.0
REIM	*Reimeria* Kociolek & Stoermer	REIM	2.5	3.0	3.5	2.5	4.5	3.0
RGBL	*Rhopalodia gibberula* (Ehrenberg) O.Müller	RHOP	2.0	3.0	3.0	3.0	4.0	3.0
RGIB	*Rhopalodia gibba* (Ehrenberg) O. Müller	RHOP	2.0	3.0	4.0	3.0	4.0	3.0
RHOI	Rhoicosphenia Grunow	RHOI	1.0	3.0	3.0	3.0	4.0	3.0
RHOP	Rhopalodia Müller	RHOP	2.0	3.0	2.8	3.0	4.0	3.0
RMUS	*Rhopalodia musculus* (Kützing) O. Müller	RHOP	2.0	3.0	2.0	3.0	4.0	3.0
ROPE	*Rhopalodia operculata* (Agardh) Hakansson	RHOP	2.0	3.0	2.0	3.0	4.0	3.0
RSIN	*Reimeria sinuata* (Gregory) Kociolek & Stoermer	REIM	3.0	3.0	4.0	2.0	5.0	3.0
RUNI	*Reimeria uniseriata* Sala, Guerrero & Ferrario	REIM	2.0	3.0	3.0	3.0	4.0	3.0
SAGA	*Stephanodiscus agassizensis* Håkansson & Kling	STEP	1.0	2.0	2.0	2.0	3.0	2.0
SANC	*Stenopterobia anceps* (Lewis) Brébisson	STEN	5.0	3.0	5.0	3.0	5.0	3.0
SANG	*Surirella angusta* Kützing	SURI	1.0	2.0	3.0	2.0	3.0	2.0
SBRE	*Surirella brebissonii* Krammer & Lange-Bertalot	SURI	2.0	2.0	3.0	2.0	3.0	2.0
SCON	*Staurosira construens* Ehrenberg	STRS	2.0	3.0	3.0	2.0	3.0	2.0
SCRU	*Surirella crumena* Brébisson	SURI	1.0	2.0	1.0	2.0	3.0	2.0
SELI	*Staurosira elliptica* (Schumann) Williams & Round	STRS	2.0	2.0	2.0	2.0	3.0	2.0
SELL	*Sellaphora* Mereschkowsky	SELL	1.7	1.3	2.3	2.0	2.0	1.7
SHAN	*Stephanodiscus hantzschii* Grunow	STEP	1.0	1.0	2.0	2.0	2.0	1.0

SIDE	*Simonsenia delognei* Lange-Bertalot	SIMO	2.0	2.0	3.0	2.0	4.0	3.0
SIMO	*Simonsenia* Lange-Bertalot	SIMO	2.0	2.0	3.0	2.0	4.0	3.0
SMNA	*Seminavis* D.G. Mann	SMNA	1.0	2.0	2.0	2.0	4.0	3.0
SMST	*Seminavis strigosa* (Hustedt) Danieledis & Economou-Amilli	SMNA	1.0	2.0	2.0	2.0	4.0	3.0
SOVI	*Surirella ovalis* Brébisson	SURI	2.0	2.0	3.0	2.0	3.0	2.0
SPHO	*Stauroneis phoenicenteron* (Nitzsch) Ehrenberg	STAU	3.0	3.0	4.0	3.0	5.0	3.0
SPIN	*Staurosirella pinnata* (Ehrenberg) Williams & Round	STRL	3.0	2.0	3.0	2.0	4.0	2.0
SPUP	*Sellaphora pupula* (Kützing) Mereschkowksy	SELL	1.0	1.0	2.0	2.0	1.0	1.0
SSEM	*Sellaphora seminulum* (Grunow) D.G. Mann	SELL	1.0	1.0	2.0	2.0	1.0	1.0
SSMI	*Stauroneis smithii* Grunow	STAU	2.0	3.0	4.0	3.0	5.0	3.0
SSTM	*Sellaphora stroemii* (Hustedt) D.G. Mann	SELL	3.0	2.0	3.0	2.0	4.0	3.0
STAN	Stauroneis anceps Ehrenberg	STAU	3.0	2.0	4.0	2.0	5.0	3.0
STAU	*Stauroneis* Ehrenberg	STAU	2.7	2.7	4.0	2.7	5.0	3.0
STDE	*Stenopterobia delicatissima* (Lewis) Brébisson	STEN	5.0	3.0	5.0	3.0	5.0	3.0
STEN	*Stenopterobia* Brébisson	STEN	5.0	3.0	5.0	3.0	5.0	3.0
STEP	*Stephanodiscus* Ehrenberg	STEP	1.0	1.7	2.0	2.0	2.7	1.7
STMI	*Stephanodiscus minutulus* (Kützing) Cleve & Moller	STEP	1.0	1.0	2.0	2.0	2.0	1.0
STRL	Staurosirella D.M. Williams & F.E. Round emend Morales	STRL	3.0	2.0	3.0	2.0	4.0	2.0
STRS	*Staurosira* (C.G. Ehrenberg) D.M. Williams &. Round	STRS	2.0	2.5	2.5	2.0	3.0	2.0
SURI	*Surirella* Turpin	SURI	1.5	2.0	2.5	2.0	3.0	2.0
TABE	*Tabellaria* Ehrenberg	TABE	5.0	3.0	5.0	3.0	5.0	3.0
TABU	*Tabularia* D.M. Williams & Round	TABU	1.0	1.0	1.0	1.0	3.0	2.0
TAPI	*Tryblionella apiculata* Gregory	TRYB	1.0	3.0	2.0	2.0	3.0	2.0
TCAL	*Tryblionella calida* (Grunow) D.G. Mann	TRYB	1.0	2.0	2.0	1.0	3.0	2.0
TCOA	*Tryblionella coarctata* (Grunow) D.G. Mann	TRYB	1.0	3.0	2.0	1.0	4.0	2.0
TDEB	*Tryblionella debilis* Arnott	TRYB	2.0	3.0	3.0	3.0	5.0	2.0

TFAS	*Tabularia fasciculata* (Agardh) D.M. Williams & Round	TABU	1.0	1.0	1.0	1.0	3.0	2.0
TFLO	*Tabellaria flocculosa* (Roth) Kützing	TABE	5.0	3.0	5.0	3.0	5.0	3.0
TGRL	*Tryblionella gracilis* W. Smith	TRYB	1.0	2.0	2.0	1.0	3.0	2.0
THAL	*Thalassiosira* P.T. Cleve	THAL	1.0	2.0	1.0	1.0	3.0	2.0
THUN	*Tryblionella hungarica* (Grunow) D.G. Mann	TRYB	1.0	3.0	1.0	2.0	3.0	2.0
TLEV	*Tryblionella levidensis* W.M. Smith	TRYB	2.0	3.0	2.0	1.0	4.0	3.0
TLIT	*Tryblionella littoralis* (Grunow) D.G. Mann	TRYB	1.0	2.0	1.0	2.0	3.0	3.0
TPSN	*Thalassiosira pseudonana* Hasle & Heimdal	THAL	1.0	2.0	1.0	1.0	3.0	2.0
TRYB	*Tryblionella* W. Smith	TRYB	1.3	2.6	1.9	1.6	3.5	2.3
TWEI	*Thalassiosira weissflogii* (Grunow) Fryxell & Hasle	THAL	1.0	2.0	1.0	1.0	3.0	2.0
UACU	*Ulnaria acus* (Kützing) Aboal	ULNA	1.0	2.0	2.0	2.0	4.0	2.0
UAMP	*Ulnaria amphirhynchus* (Ehrenberg) Compère & Bukhtiyarova	ULNA	1.0	2.0	2.0	2.0	4.0	2.0
ULNA	*Ulnaria* Compère	ULNA	1.8	2.0	2.3	2.0	4.0	2.0
UULN	*Ulnaria ulna* (Nitzsch) Compère	ULNA	2.0	2.0	2.0	2.0	4.0	2.0
UUNG	*Ulnaria ungeriana* (Grunow) Compère	ULNA	3.0	2.0	3.0	2.0	4.0	2.0

Table A.2: Species autecological description inferred from expert knowledge.

Code	**Denomination**	**Genus**	**General ecology of species (All criteria included)**
AAMB	*Aulacoseira ambigua* (Grunow) Simonsen	AULA	This taxon is typically found in the plankton of meso-eutrophic lakes and rivers.
AAMC	*Aulacoseira ambigua* fo. *curvata* Skabicevsky	AULA	Variation of nominate species, cells curved in presence of high nutrient load.
ABRY	*Adlafia bryophila* (Petersen) Moser, Lange-Bertalot & Metzeltin	ADLF	Aerophytic species occurring on intermittently wet bryophytes. Ecology uncertain.
ACAF	*Achnanthidium affine* (Grunow) Czarnecki	ACHD	Pioneer, apically attached. Found in clean calcareous water with moderate electrolyte content. Well oxygenated water.
ACHD	*Achnanthidium* Kützing	ACHD	
ACHN	*Achnanthes* Bory de St. Vincent	ACHN	
ACOF	*Amphora coffeaeformis* (Agardh) Kützing	AMPH	This taxon is found in brackish and saline waters with high electrolyte content.

ACOP	*Amphora copulata* (Kützing) Schoeman & Archibald	AMPH	This taxon is found in waters with moderate electrolyte content, sometimes extending into brackish conditions.
ADCR	*Achnanthidium crassum* (Hustedt) Potapova & Ponader (*Achnanthes crassa*)	ACHD	This taxon is found in slow moving alkaline streams.
ADEG	*Achnanthidium exiguum* (Grunow) Czarnecki (*Gogorevia exilis*)	ACHD	This taxon has a wide ecological amplitude and is tolerant of Low light, high temperature conditions.
ADEU	*Achnanthidium eutrophilum* (Lange-Bertalot) Lange-Bertalot	ACHD	Pioneer taxon found in oxygenated eutrophic waters with low to moderate pollution
ADLF	*Adlafia* Moser, Lange-Bertalot & Metzeltin	ADLF	
ADMA	*Achnanthidium macrocephalum* (Hustedt) Round & Bukhtiyarova	ACHD	Oligotrophic, circumnutral to acidic water. Low organic material.
ADMI	*Achnanthidium minutissimum* (Kützing) Czarnecki	ACHD	Found in well oxygenated water (Species complex) - often leading to misattribution to ecological requirements. Pioneer.
ADRI	*Achnanthidium rivulare* Potapova & Ponader	ACHD	This taxon occurs in oligo- to mesotrophic circumneutral waters.
ADSA	*Achnanthidium saprophilum* (Kobayasi & Mayama) Round & Bukhtiyarova	ACHD	Found in organically enriched eutrophic fresh waters
ADST	*Achnathidium standeri* (Cholnoky) J.C.Taylor, E.Morales & L.Ector (*Achnanthes standeri*)	ACHN	Endemic to South Africa. Indicator of clean oligotrophic water with low electrolyte content
AEXG	*Achnanthes exigua* Kützing *(Gogorevia exilis* (Kützing) Kulikovskiy and Kociolek)	ACHN	Can tolerate high temperatures of 40 degrees Celsius
AFON	*Amphora fontinalis* Hustedt	AMPH	A rare taxon, possibly occurring in acidic waters.
AFOR	*Asterionella formosa* Hassall	ASTE	Strictly planktonic species occurring in eutrophic waters, not tolerant to organic pollution, causes taste and odour problems in water
AINA	*Amphora inariensis* Krammer	AMPH	This taxon occurs in oligotrophic waters with moderate electrolyte conditions.
ALIO	*Achnanthidium linearioides* Lange-Bertalot	ACHN	Found in well oxygenated slightly acidic oligotrophic water
AMMO	*Amphora montana* Krasske	AMPH	A rare taxon found in alkaline waters. Rarely dominant.
AMPH	*Amphora* Ehrenberg	AMPH	
AMPI	*Amphipleura* Kützing	AMPI	
AMUZ	*Aulacoseira muzzanensis* (Meister) Krammer	AULA	This taxon is typically found in the plankton and benthos of eutrophic lakes and rivers
ANOM	*Anomoeneis* Pfitzer	ANOM	
ANOR	*Amphora normanii* Rabenhorst	AMPH	This aerophilic taxon occurs in mountainous areas and wetland biotopes.

AOBG	Achnanhes oblongella Østrup (*Platessa oblongella* (Østrup) C.E.Wetzel, Lange-Bertalot & Ector)	ACHN	This taxon is more often found in small, circumneutral, oligotrophic streams with low electrolyte content.
AOVA	*Amphora ovalis* (Kützing) Kützing	AMPH	This taxon is found in waters with moderate electrolyte content, extending into brackish and saline conditions.
APED	*Almphora pediculus* (Kützing) Grunow	AMPH	This taxon occurs in waters with moderate electrolyte content and tolerates critical levels of pollution. Often epiphytic on other algae, including diatoms.
APEL	*Amphipleura pellucida* Kützing	AMPI	This taxon is found in alkaline waters of moderate to high electrolyte content, extending in brackish conditions
ASAF	*Achnanthes subaffinis* Cholnoky	ACHN	Found in slow moving oligotrophic streams. Western cape acidic water
ASCO	*Anomoeoneis sphaerophora* fo. *costata* (Kützing)Schmidt	ANOM	This taxon occurs in inland saline waters with high electrolyte content. Tolerant to critical levels of pollution.
ASPH	*Anomoeoneis sphaerophora* (Ehrenberg) Pfitzer	ANOM	This taxon occurs in the littoral zone of waters with moderate to high electrolyte content, extending into brackish conditions and coastal waters. Tolerant to critical levels of pollution.
ASTE	*Asterionella* A.H. Hassall	ASTE	
ASWA	*Achnanthes swazi* Cholnoky	ACHN	Endemic to South Africa, found in oligotrophic, circumneutral waters. Low organic material.
AUCS	*Aulacoseira crassipunctata* Krammer	AULA	This taxon is typically found in the plankton and benthos of eutrophic lakes and rivers
AUGA	*Aulacoseira granulata* var. *angustissima* (O.Müller) Simonsen	AULA	This taxon is typically found in the plankton and benthos of eutrophic lakes and rivers
AUGR	*Aulacoseira granulata* (Ehrenberg) Simonsen	AULA	This taxon is typically found in the plankton and benthos of eutrophic lakes and rivers
AULA	*Aulacoseira* Thwaites	AULA	
AUSB	*Aulacoseira subborealis* (Nygaard) Denys, Muylaert & Krammer	AULA	Planktonic taxon in alkaline eutrophic lakes and rivers with moderate electrolyte content
AVEN	*Amphora veneta* Kützing	AMPH	This taxon is found in waters with elevated electrolyte content. Tolerant to critical and very heavy levels of pollution.
BACI	*Bacillaria* Gmelin	BACI	
BBRE	*Brachysira brebissonii* Ross	BRAC	This taxon is found in acidic, oligotrophic, electrolyte-poor waters. A good indicator for naturally acidic water with no anthropogenic impacts.

BNEO	*Brachysira neoexilis* Lange-Bertalot	BRAC	This taxon occurs in clean oligo-mesotrophic waters. Occurs in electrolyte poor, acidic water as well as calcareous, alkaline biotopes.
BPAR	*Bacillaria paradoxa* Gmelin	BACI	A cosmopolitan species widespread in very electrolyte-rich and brackish waters particularly near the coast. The cells form a unique type of motile colony, in which individual cells slide to and fro with respect to each other. They are held together by interlocking grooves on the raphe-sterna. The colony extends to form a linear array, where only the poles of the cells are touching then retracts to form a tabular array, with all the cells side by side.
BRAC	*Brachysira* Kützing	BRAC	
BVIT	*Brachysira vitrea* (Grunow) Ross	BRAC	This taxon occurs in clean oligo-mesotrophic waters. Occurs in electrolyte poor, acidic water as well as calcareous, alkaline biotopes.
BWYG	*Brachysira wygaschii* Lange-Bertalot	BRAC	This taxon is found in oligotrophic, electrolyte poor waters.
BZEL	*Brachysira zellensis* (Grunow) Round & Mann	BRAC	This taxon is found in oligotrophic waters with low electrolyte content.
CACD	*Craticula acidoclinata* Lange-Bertalot & Metzeltin	CRAT	Occurring in waters with low organic material and low electrolyte content
CACM	*Craticula accomodiformis* Lange-Bertalot	CRAT	This taxon is found in eutrophic electrolyte rich waters with low to moderate levels of pollution. Tropic to subtropic species
CALO	*Caloneis* Cleve	CALO	
CAMB	*Craticula ambigua* (Ehrenberg) D.G. Mann	CRAT	This taxon occurs in moderate to electrolyte rich eutrophic waters. Resistant to critical levels of pollution
CAPA	*Capartogramma* Kufferath	CAPA	
CAQT	*Caloneis aequatorialis* Hustedt	CALO	This tropical to sub-tropical taxon is found in alkaline waters of South Africa.
CASP	*Cymbella aspera* (Ehrenberg) H. Peragallo	CYMB	This taxon is attached to the substratum by apically attached mucilage stalks. Found in oligotrophic waters with moderate electrolyte content.
CATO	*Cyclotella atomus* Hustedt	CYCL	Planktonic and benthic taxon found in impacted electrolyte rich streams
CBAC	*Caloneis bacillum* (Grunow) Cleve	CALO	This taxon is found in the littoral zone of waters with moderate electrolyte content. Often found on damp bryophytes.

CBAM	*Cymbopleura amphicephala* Krammer	CBPL	This taxon is found in oligo- to mesotrophic waters with low to moderate electrolyte content.
CBNA	*Cymbopleura naviculiformis* (Auerswald) Krammer	CBPL	This taxon is found in oligo- to mesotrophic waters with low to moderate electrolyte content.
CBPL	*Cymbopleura* (Krammer) Krammer	CBPL	
CCLY	*Campylodiscus clypeus* Ehrenberg	CPLD	This taxon is found in saline waters, especially in coastal regions. Also found in inland salt pans.
CCRU	*Capartogramma crucicula*(Grunow) Ross	CAPA	The ecology of this taxon is uncertain. This taxon is found in tropical and subtropical climates and has been increasingly found in South Africa in recent years.
CCRY	*Cyclotella cryptica* Reimann & Volcani	CYCL	Planktonic and benthic in impacted streams
CCST	*Cyclostephanos* Round	CCST	
CCYM	*Cymbella cymbiformis* Agardh	CYMB	This taxon is found in the littoral zone of lakes and streams and also found in small watercourses and puddles. Common in oligotrophic waters with low to very low electrolyte content.
CDUB	*Cyclostephanos dubius* (Fricke) Round	CCST	Euplanktonic species found in highly impacted water - High alkaline (chloride & calcareous)
CENG	*Cocconeis engelbrechtii* Cholnoky		Endemic to South Africa. Found in brackish inland waters (alkaline high electrolyte content)
CHAL	*Craticula halophila* (Grunow) D.G. Mann	CRAT	This taxon occurs in salt springs and water with high to very high electrolyte content
CHYA	*Caloneis hyalina* Hustedt	CALO	This aerophilic species is mostly found in the tropics and sub-tropics.
CINV	*Cyclostephanos invisitatus* (Hohn & Hellerman) Theriot Stoermer & Håkansson	CCST	Highly impacted water - High alkaline (chloride & calcareous)
CKOL	*Cymbella kolbei* Hustedt	CYMB	This taxon is found in oligotrophic alkaline waters.
CKPP	*Cymbella kappii* (Cholnoky) Cholnoky	CYMB	Distributed and very common throughout South Africa with a limited distribution in Europe and other parts of the world. Found in weakly alkaline, oligo- to mesotrophic waters with low to moderate electrolyte content.
CMED	*Cyclotella meduanae* Germain	CYCL	This taxon is found in the benthos and plankton of eutrophic, electrolyte rich lakes and rivers

CMEN	*Cyclotella meneghiniana* Kützing	CYCL	This taxon is found in the benthos and plankton of eutrophic, electrolyte rich lakes and rivers
CMLF	*Craticula molestiformis* (Hustedt) Lange-Bertalot	CRAT	Found in highly polluted waters with high nutrient and organic content
CMOL	*Caloneis molaris* (Grunow) Krammer	CALO	The ecology of this taxon is uncertain.
CNCI	*Cymbella neocistula* Krammer	CYMB	This epiphytic and epilithic taxon is found in circumneutral to slightly alkaline, mesotrophic waters with moderate to high electrolyte content.
COCE	*Cyclotella ocellata* Pantocsek	CYCL	Planktonic taxon occurring in meso-eutrophic waters with elevated pH
COCO	*Cocconeis* Ehrenberg	COCO	
CPED	*Cocconeis pediculus* Ehrenberg	COCO	Common epiphytic species occurring in waters with moderate to high electrolyte content extending into brackish conditions
CPLA	*Cocconeis placentula* Ehrenberg	COCO	Found in meso- to eutrophic flowing or standing waters. Abundant on plants and plant debris
CPLD	*Campylodiscus* Ehrenberg	CPLD	
CPLE	*Cocconeis placentula* Ehrenberg var. *euglypta* (Ehrenberg) Grunow	COCO	Found in meso- to eutrophic flowing or standing waters. Abundant on plants and plant debris
CPLI	*Cocconeis placentula* Ehrenberg var. *lineata* (Ehrenberg) Van Heurck	COCO	Found in oligo- to mesotrophic flowing or standing waters. Abundant on plants and plant debris
CRAC	*Craticula accomoda* (Hustedt) Mann	CRAT	Occurring in waters with high levels of organic and nutrient content. Indicator of pollution.
CRAT	*Craticula* Grunow	CRAT	
CRBU	*Craticula buderi* (Hustedt) Lange-Bertalot	CRAT	Dominant in platinum mine drains. Occurring in calcareous rich waters
CRCU	*Craticula cuspidata* (Kützing) D.G. Mann	CRAT	This taxon occurs in eutrophic waters with moderate to high electrolyte content, extending into brackish conditions. May tolerate critical to heavy pollution
CSAP	*Cymatopleura solea* var. *apiculata* (W. Smith) Ralfs	CYMA	This taxon occurs in eutrophic waters with moderate to high electrolyte content sometimes found in brackish conditions. Favours alkaline waters. Epipelic and epiphytic taxon in the littoral zone. Rarely dominant, sediment rich environment.
CSBM	*Craticula submolesta* (Hustedt) Lange-Bertalot	CRAT	This taxon occurs in oligotrophic circumneutral waters wit low electrolyte content.
CSHU	*Caloneis schumanniana* (Grunow) Cleve	CALO	This taxon is found in the littoral zone of oligotrophic waters with moderate

			electrolyte content. Also occurs in calcareous streams.
CSIL	*Caloneis silicula* (Ehrenberg) Cleve	CALO	This taxon occurs in the littoral zone of waters with moderate electrolyte content.
CSLP	*Cymbella subleptoceros* Krammer	CYMB	This cosmopolitan taxon in found in mesotrophic waters with moderate electrolyte content.
CSMO	*Cymbella simonsenii* Krammer	CYMB	This taxon is found in oligotrophic, calcareous waters with moderate electrolyte content.
CSOL	*Cymatopleura solea* (Brébisson) W. Smith	CYMA	This taxon occurs in eutrophic waters with moderate to high electrolyte content sometimes found in brackish conditions. Favours alkaline waters. Epipelic and epiphytic taxon in the littoral zone. Rarely dominant, sediment rich environment.
CSTT	*Discostella stelligera* (Cleve and Grunow) Houk & Klee	CYCL	Planktonic taxon found in inland rivers and lakes
CTGL	*Cymbella turgidula* Grunow	CYMB	This taxon is found in oligo- to mesotrophic, alkaline waters with moderate electrolyte content.
CTNP	Ctenophora (A. Grunow) D.M. Williams & Round	CTNP	
CTPU	*Ctenophora pulchella* (Ralfs) Williams & Round	CTNP	This taxon occurs in strongly polluted water - industrial and mining
CTUM	*Cymbella tumida* (Brébisson) Van Heurck	CYMB	This taxon occurs in the littoral zone of standing and flowing waters. Found in oligo- to mesotrophic waters with moderate electrolyte content.
CVIX	*Craticula vixnegligenda* Lange-Bertalot	CRAT	Clean water, low organic content
CWOL	*Cyclotella woltereckii* Hustedt	CYCL	Planktonic. Possibly a small form of *Discostella pseudostelligera*
CYCL	*Cyclotella* Kützing	CYCL	
CYMA	*Cymatopleura* W. Smith	CYMA	
CYMB	*Cymbella* C.Agardh	CYMB	
DCOF	*Diadesmis confervacea* Kützing	DIAM	This taxon is found in a wide range of waters including eutrophic, electrolyte rich and extremely polluted waters
DENT	*Denticula* Kützing	DENT	
DIAM	*Diadesmis* Kützing	DIAM	
DIAT	*Diatoma* Bory de St. Vincent	DIAT	
DIPL	*Diploneis* Ehrenberg	DIPL	
DISC	*Discostella* Houk & Klee	DISC	
DKUE	*Denticula kuetzingii* Grunow	DENT	This taxon occurs in waters with moderate to high electrolyte content.
DPST	*Discostella pseudostelligera* (Hustedt) Houk & Klee	DISC	Found in the plankton of freshwater inland lakes and rivers
DSMI	*Diploneis smithii* (Brébisson) Cleve	DIPL	Taxon found in brackish to weakly saline inland waters

DSTE	*Discostella stelligera* (Cleve & Grunow) Houk & Klee	DISC	Found in the plankton of freshwater inland lakes and rivers
DSUB	*Denticula subtilis* Grunow	DENT	This taxon occurs in electrolyte-rich and brackish waters.
DSUN	*Denticula sundayensis* Archibald	DENT	This taxon occurs in electrolyte-rich and brackish waters.
DVUL	*Diatoma vulgaris* Bory	DIAT	This taxon forms blooms during winter in waters with moderate electrolyte content. Not tolerant of elevated organic content. Occurs in meso-eutrophic waters
DWOL	*Discostella woltereckii* (Hustedt) Houk & Klee	DISC	Found in the plankton of freshwater inland lakes and rivers under silica limited conditions
EADN	*Epithemia adnata* (Kützing) Brébisson	EPIT	This taxon occurs in standing and flowing waters with moderate to high electrolyte content, also extending into brackish conditions. Often contains nitrogen fixing endosymbiotic cyanobacteria, adapted to warm water.
EBIL	*Eunotia bilunaris* (Ehrenberg) Mills	EUNO	This taxon is an indicator of acidic water with low electrolyte content
ECAE	*Encyonema caespitosum* Kützing	ENCY	This taxon is found in oligo- to eutrophic waters favouring those with a high electrolyte content. Tolerating critical levels of pollution. Cells found within a mucilage tube.
ECBU	*Encyonopsis buedelii* Krammer	ENCP	This taxon was described from the Golden Gate National Park. Found in Oligotrophic, slightly acidic waters on the Mont-Aux-Sources Plateau (3000m)
ECES	*Encyonopsis cesatii* (Rabenhorst) Krammer	ENCP	This montane taxon occurs in well oxygenated biotopes e.g. rock faces, mosses, springs and streams.
ECFA	*Encyonopsis falaisensis* (Grunow) Krammer	ENCP	This taxon occurs in oligotrophic, oxygen rich waters with low to moderate electrolyte content.
ECKR	*Encyonopsis krammeri* Reichardt	ENCP	This taxon is found in oligotrophic, slightly acidic waters with low electrolyte content.
ECPM	*Encyonopsis minuta* Krammer & Reichardt	ENCP	This taxon occurs in calcareous waters with moderate electrolyte content. Requires a well-oxygenated environment.
EEXI	*Eunotia exigua* (Brébisson) Rabenhorst	EUNO	This taxon is found is oligotrophic, nutrient poor water with high acidity
EFLE	*Eunotia flexuosa* (Brébisson) Kützing	EUNO	This taxon occurs in oligotrophic acidic waters
EFOR	*Eunotia formica* Ehrenberg	EUNO	This taxon occurs is dystrophic environments and is commonly found in the western cape, rarely dominant.

EINC	*Eunotia incisa* Gregory	EUNO	This taxon occurs is upland streams in oligotrophic acidic, electrolyte-poor waters
EMIN	*Eunotia minor* (Kützing) Grunow	EUNO	This taxon occurs in circumneutral to slightly acidic rivers, pools and springs
ENCM	*Encyonopsis microcephala* (Grunow) Krammer	ENCP	This taxon occurs in calcareous waters with moderate electrolyte content.
ENCP	Encyonopsis Krammer	ENCP	
ENCY	*Encyonema* Kützing	ENCY	
ENLS	*Encyonopsis leei* Krammer	ENCP	This taxon has become widespread in South Africa in recent years. Found in slightly acidic, oligo- to mesotrophic waters with low to moderate electrolyte content.
ENME	*Encyonema mesianum* (Cholnoky) D.G. Mann	ENCY	This montane taxon occurs in weakly acidic waters.
ENMI	*Encyonema minutum* (Hilse) D.G. Mann	ENCY	This taxon is found in oligotrophic waters with moderate electrolyte content.
ENNG	*Encyonema neogracile* Krammer	ENCY	This taxon occurs in oligotrophic, electrolyte-poor waters.
ENVE	*Encyonema ventricosum* (Agardh) Grunow	ENCY	This taxon occurs in well-oxygenated, alkaline waters.
EOAR	*Eolimna archibaldii* Taylor & Lange-Bertalot (*Sellaphora archibaldii* (Taylor & Lange-Bertalot) Ács, C.E. Wetzel & Ector)	EOLI	This taxon is endemic to South Africa and occurs in alkaline eutrophic waters with elevated electrolyte content
EOLI	*Eolimna* Lange-Bertalot & Schiller	EOLI	
EOMI	*Eolimna minima* (Grunow) Lange-Bertalot (*Sellaphora atomoides*)	EOLI	This taxon occurs in a wide range of waters including heavily polluted biotopes. Commonly associated with detritus.
EPIT	*Epithemia* Kützing	EPIT	
EPUN	*Eunotia pectinalis* var. *undulata* (Ralfs) Rabenhorst	EUNO	This taxon is found is weakly acidic to circumneutral waters with low electrolyte content
ERAY	*Encyonopsis raytonensis* (Cholnoky) Krammer	ENCP	This taxon has only been recorded from South Africa, found in acidic, well oxygenated waters.
ERHO	*Eunotia rhomboidea* Hustedt	EUNO	This taxon is an indicator of clean pollutant free water
ESBM	*Eolimna subminuscula* (Manguin) Moser, Lange-Bertalot & Metzeltin (*Craticula subminiscula*)	EOLI	This taxon occurs in electrolyte rich rivers with high levels of pollution
ESLE	*Encyonema silesiacum* (Bleisch) D.G. Mann	ENCY	This taxon occurs in standing and flowing oligo- to eutrophic waters. May tolerate strongly polluted conditions.
ESOR	*Epithemia sorex* Kützing	EPIT	This taxon is found in flowing and standing waters with moderate to high electrolyte content. Often contains nitrogen fixing endosymbiotic cyanobacteria, warm water adapted.

ESUM	*Encyonopsis subminuta* Krammer & Reichardt	ENCP	This taxon occurs in calcareous waters with moderate electrolyte content. Requires a well-oxygenated environment.
ETEN	*Eunotia tenella* (Grunow)Hustedt	EUNO	This taxon is found in highly acidic environment but not as acidic as nominate variety (*Eunotia exigua*)
EUNO	*Eunotia* Ehrenberg	EUNO	
FALL	*Fallacia* Stickle & D.G. Mann	FALL	
FBCP	*Fragilaria biceps* (Kützing) Lange-Bertalot (*Ulnaria biceps*)	FRAG	This taxon is found alongside Fragilaria ulna in the benthos of meso-eutrophic waters
FCAP	*Fragilaria capucina* Desmazieres	FRAG	*Fragilaria capucina* species complex. Lower organic material with moderate electrolyte content in oligo-meso trophic waters
FCRO	*Fragilaria crotonensis* Kitton	FRAG	Strictly planktonic taxon occurring in slightly alkaline oligo- eutrophic waters with low organic content,
FCRS	*Frustulia crassinervia* (Brébisson) Lange-Bertalot & Krammer	FRUS	This taxon occurs in oligotrophic standing waters with low electrolyte content
FCRU	*Fragilaria capucina* var. *rumpens* (Kützing) Lange-Bertalot (*Fragilaria rumpens*)	FRAG	Taxon found in the benthos of oligo-mesotrophic waters
FCVA	*Fragilaria capucina* Desmazieres var. *vaucheriae* (Kützing) Lange-Bertalot (*Fragilaria vaucheriae*)	FRAG	No clearly defined distribution
FINS	*Fallacia insociabilis* (Krasske) D.G. Mann	FALL	This aerophylic taxon is found in a wide range of waters
FITU	*Fistulifera* Lange-Bertalot	FITU	
FMGL	*Frustulia magaliesmontana* Cholnoky	FRUS	Ecological amplitude not precisely known. Thought to occur in acidic oligotrophic waters with low electrolyte content
FMOC	*Fallacia monoculata* (Hustedt) D.G. Mann	FALL	This taxon occurs in a wide range of waters with moderate to high electrolyte content
FNAN	*Fragilaria nanana* Lange-Bertalot	FRAG	Planktonic taxon found in oligotrophic waters with low organic material
FPMG	*Frustulia pseudomagaliesmontana* Camburn & Charles	FRUS	This taxon occurs in circumneitral to acidic waters. Oligotrophic waters with low electrolyte content.
FPSC	*Fragilaria parasitica* var. *subconstricta* Grunow	FRAG	A benthic taxon with a low occurrence, usually attached to other species occurring in meso- eutrophic circumneutral waters
FPYG	*Fallacia pygmaea* (Kützing) Stickle & D.G. Mann	FALL	This taxon occurs in waters with elevated electrolyte content and is tolerant to critical levels of pollution
FRAG	*Fragilaria* Lyngbye	FRAG	

FROS	*Frustulia rostrata* Hustedt	FRUS	This taxon is associated with bryophytes and usually occurs in acidic standing or flowing waters.
FRUM	*Fragilaria rumpens* (Kützing) G.W.F.Carlson	FRAG	This taxon occurs in oligotrophic circumneutral waters with low electrolyte content.
FRUS	*Frustulia* Rabenhorst	FRUS	
FSAP	*Fistulifera saprophila* (Lange-Bertalot & Bonik) Lange-Bertalot	FITU	This species is highly resistant to pollution. Occurs in eutrophic anthropogenically impacted waters
FSAX	*Frustulia saxonica* Rabenhorst	FRUS	Indicator for low organic content. Occurring in dystrophic acidic, electrolyte poor waters
FSBH	*Fallacia subhamulata* (Grunow) D.G. Mann	FALL	This cosmopolitan taxon occurs across a wide range of waters
FTEN	*Fragilaria tenera* (W. Smith) Lange-Bertalot	FRAG	Planktonic species found in meso-eutrophic waters with low organic material.
FTNR	*Fallacia tenera* (Hustedt) Mann in Round	FALL	This taxon occurs in waters with high to very high electrolyte content (brackish conditions)
FTUG	*Frustulia tugelae* Cholnoky	FRUS	A rare endemic species of South Africa. Occurring in the weakly acidic waters of the eastern mountainous regions of South Africa. Oligtrophic waters with low electrolyte content.
FUAC	*Fragilaria ulna* var. *acus* (Kützing) Lange-Bertalot (*Ulnaria acus*)	FRAG	Taxon occurring in the benthos of meso-eutrophic alkaline lakes and rivers
FUMP	*Fallacia umpatica* (Cholnoky) D.G. Mann	FALL	A South African species described from brackish waters
FVAU	*Fragilaria vaucheriae* (Kützing) Petersen	FRAG	Wide ecological range. Meso-Eutrophic preference
FVUL	*Frustulia vulgaris* (Thwaites) De Toni	FRUS	This taxon has a very wide ecological amplitude, occurring in fresh to brackish waters also ranging from oligotrophic to highly polluted waters
FWEI	*Frustulia weinholdii* Hustedt	FRUS	The ecological amplitude is not well known. Thought to occur in oligo-eutrophic waters with low to moderate electrolyte content
GACU	*Gomphonema acuminatum* Ehrenberg	GOMP	This taxon is found in circumneutral to weakly alkaline waters. Tolerant to slight or moderate pollution. Attached to substratum by means of a mucilage stalk.
GAFF	*Gomphonema affine* Kützing	GOMP	This tropical/sub-tropical species is found in waters with elevated electrolyte content.
GANG	*Gomphonema angustatum* (Kützing) Rabenhorst	GOMP	This taxon is common in oligotrophic waters. Found over a range of pH and

			electrolyte concentrations, including calcium rich waters.
GCAP	*Gomphonema capitatum* Ehrenberg	GOMP	This taxon is found in oligotrophic waters with elevated electrolyte content. Not tolerant of more than moderate levels of pollution.
GCLA	*Gomphonema clavatum* Ehrenberg	GOMP	This montane taxon occurs in oligotrophic waters but can tolerate high electrolyte content.
GDEC	*Geissleria decussis* (Østrup) Lange-Bertalot & Metzeltin	GEIS	This taxon is found in eutrophic, unpolluted or moderately polluted waters with slightly elevated electrolyte content
GEIS	*Geissleria* Lange-Bertalot & Metzeltin	GEIS	
GEXL	*Gomphonema exilissimum* (Grunow) Lange-Bertalot & Reichardt	GOMP	Delineated form of *Gomphonema parvulum*.
GGRA	*Gomphonema gracile* Ehrenberg	GOMP	This taxon is found in electrolyte rich environments but not tolerant of more than moderate levels of pollution.
GINS	*Gomphonema insigne* Gregory	GOMP	This taxon is found in electrolyte-rich waters.
GITA	*Gomphonema italicum* Kützing	GOMP	This taxon is found in oligotrophic to slightly eutrophic waters with elevated electrolyte content. Not tolerant of more than moderate levels of pollution.
GLTC	*Gomphonema laticollum* Reichardt	GOMP	This subcosmopolitan taxon is found in slightly eutrophic habitats. Attached to the substratum by means of dichotomous mucilage stalks.
GMIN	Gomphonema minutum (Agardh) Agardh	GOMP	This taxon is found in eutrophic waters. Not tolerant of more than moderate levels of pollution. Attached to substratum by means of dichotomous mucilage stalks.
GOMP	*Gomphonema* Ehrenberg	GOMP	
GPAR	*Gomphonema parvulum* (Kützing) Kützing	GOMP	This taxon has a very wide ecological amplitude, occurring in a range of waters. Tolerant to extreme levels of pollution.
GPAS	*Gomphonema parvulum* var. *parvulum* f. *saprophilum* Lange-Bert.&Reichardt	GOMP	This taxon has a very wide ecological amplitude, occurring in a range of waters. Tolerant to extreme levels of pollution.
GPLA	*Gomphonema parvulum* var. *lagenula* (Kützing) Frenguelli	GOMP	Little is known of this taxon's ecology. A poorly delineated form of *Gomphonema parvulum*.
GPPA	*Gomphonema parvulum* var. *parvulius* Lange-Bertalot & Reichardt	GOMP	This taxon is found in circumneutal, oligotrophic, electrolyte-poor waters.
GPRI	*Gomphonema pumilum* var. *rigidum* Reichardt & Lange-Bertalot	GOMP	This taxon is found in meso- to eutrophic waters with moderate electrolyte content. Not tolerant of more than critical levels of pollution.

GPSA	*Gomphonema pseudoaugur* Lange-Bertalot	GOMP	This taxon is found in meso- to eutrophic waters, not tolerating more than critical levels of pollution. Attached to substratum by means of dichotomous mucilage stalks.
GPUM	*Gomphonema pumilum* (Grunow) Reichardt & Lange-Bertalot	GOMP	This taxon is found in meso- to eutrophic waters with moderate electrolyte content. Not tolerant of more than moderate levels of pollution.
GRAU	*Gyrosigma rautenbachiae* Cholnoky	GYRO	This taxon occurs in slow moving brackish water and is tolerant of industrial pollution
GSCA	*Gyrosigma scalproides* (Rabenhorst)Cleve	GYRO	This taxon occurs in highly turbid waters with moderate to high electrolyte content
GTRU	*Gomphonema truncatum* Ehrenberg	GOMP	This taxon is found in oligotrophic waters with elevated electrolyte content. Not tolerant of more than moderate levels of pollution.
GVNU	*Gomphonema venusta* Passy, Kociolek & Lowe	GOMP	This taxon was described from South Africa and occurs very commonly in the northern and central parts of the country. Found in circumneutral to weakly alkaline, oligo- to mesotrophic waters with a low to moderate electrolyte content.
GYAC	*Gyrosigma acuminatum* (Kützing) Rabenhorst	GYRO	This taxon is found in electrolyte rich to brackish waters and can tolerate high levels of organic pollution
GYAT	*Gyrosigma attenuatum* (Kützing) Rabenhorst	GYRO	Occurring in sediment rich water, hghly motile. Occurs in moderate to high electrolyte content but unable to tolerate high levels of pollution
GYRO	*Gyrosigma* Hassall	GYRO	
HACO	*Halamphora coffeaeformis* (Agardh) Levkov	HALA	This taxon is found in brackish and saline waters with high electrolyte content.
HALA	*Halamphora* (Cleve) Levkov	HALA	
HANT	*Hantzschia* Grunow	HANT	
HAVT	*Hippodonta avittata* (Cholnoky) Lange-Bertalot, Metzeltin & Witkowski	HIPO	The ecology of this taxon is uncertain. However, it is often found en masse in waters with high electrolyte content.
HCAP	*Hippodonta capitata* (Ehrenberg) Lange-Bertalot, Metzeltin & Witkowski	HIPO	This taxon occurs is eutrophic waters with moderate to high electrolyte content, occasionally brackish. Tolerant of critical pollution levels
HDIS	*Hantzschia distinctepunctata* Hustedt	HANT	This taxon occurs in waters with very high electrolyte content, extending into brackish conditions.
HIPO	*Hippodonta* Lange-Bertalot, Metzeltin & Witkowski	HIPO	

HNOR	*Halamphora normanii* (Rabenhorst) Levkov	HALA	This aerophilic taxon occurs in mountainous areas and wetland biotopes.
KOBL	*Kobayasiella* Lange-Bertalot	KOBL	
KOSU	*Kobayasiella subtilissima* (Cleve) Lange-Bertalot	KOBL	A cosmopolitan species occurring in acidic, electrolyte rich waters
LEMN	*Lemnicola* Round & Basson	LEMN	
LHUN	*Lemnicola hungarica* (Grunow) Round & Basson	LEMN	This epiphytic taxon is mostly associated with *Lemna* spp. (Duckweed) and occurs in weakly alkaline waters with moderate to elevated electrolyte content. May also occur in critically polluted waters.
MAAT	*Mayamaea atomus* (Kützing) Lange-Bertalot	MAYA	Aerophilous species, found in alkaline, heavily polluted waters with a high electrolyte content. One of the most pollution resistant diatoms, but also occurring in moderate quality waters probably associated with a micro-habitat e.g. organic detritus
MAPE	*Mayamaea atomus* var. *permitis* (Hustedt) Lange-Bertalot	MAYA	Aerophilous species, found in alkaline, heavily polluted waters with a high electrolyte content. One of the most pollution resistant diatoms, but also occurring in moderate quality waters probably associated with a micro-habitat e.g. organic detritus
MAST	*Mastogloia* Thwaites	MAST	
MAYA	*Mayamaea* Lange-Bertalot	MAYA	
MDAN	*Mastogloia dansei* (Thwaites) W.M.Smith	MAST	This epiphytic or epipelic taxon occurs in waters with moderate to high electrolyte content. Rarely abundant.
MELL	*Mastogloia elliptica* (C.A. Agardh) Cleve	MAST	This taxon is a cosmopolitan brackish water species
MELO	*Melosira* Agardh	MELO	
MSMI	*Mastogloia smithii* Thwaites	MAST	This species is commonly found in brackish waters but also occurs in waters with moderate to high electrolyte content
MVAR	*Melosira varians* Agardh	MELO	This taxon typically occurs in the plankton and benthos of lakes and rivers with high nutrients, lower organic content and moderate electrolyte content.
NAAN	*Navicula angusta* Grunow	NAVI	This taxon is a good indicator for acidic, oligotrophic, un-impacted, electrolyte poor waters.
NACD	*Nitzschia acidoclinata* Lange-Bertalot	NITZ	This taxon is found in acidic, oligotrophic, electrolyte poor, small bodies of water.
NACI	*Nitzschia acicularis* (Kützing) W.M.Smith	NITZ	This planktonic and epipelic taxon is found in eutrophic waters with moderate to high electrolyte content. Tolerant of

			strongly polluted conditions but not of extremely polluted conditions.
NAGN	*Nitzschia agnita* Hustedt	NITZ	This taxon is found in electrolyte-rich and brackish waters.
NAGW	*Nitzschia agnewii* Cholnoky	NITZ	This taxon occurs in eutrophic waters in South Africa.
NAMA	*Navicula arvensis* Hustedt var. *maior* Lange-Bertalot	NAVI	This taxon is found in waters with moderate to high electrolyte content
NAMC	*Nitzschia amplectens* Hustedt	NITZ	Intertidal species
NAMP	*Nitzschia amphibia* Grunow	NITZ	This taxon in found in eutrophic waters over a range from electrolyte-poor to electrolyte-rich waters.
NANT	*Navicula antonii* Lange-Bertalot	NAVI	This taxon is found in eutrophic to hypereutrophic waters with moderate to high electrolyte content. Tolerant to strongly polluted conditions. A good indicator for such anthropogenically impacted waters.
NAUR	*Nitzschia aurariae* Cholnoky	NITZ	This taxon occurs in electrolyte-rich waters and sporadically is other types of water. Common in gold-mine effluents.
NAVG	*Navigiolum* Lange-Bertalot, Cavacini. Tagliaventi & Alfinito	NAVG	
NAVI	*Navicula* Bory de St. Vincent	NAVI	
NBCL	*Nitzschia bacillum* Hustedt	NITZ	This taxon occurs in oligotrophic, electrolyte-rich waters.
NCIN	*Navicula cincta* (Ehrenberg) Ralfs	NAVI	This taxon is found in calcareous, oligotrophic waters.
NCLA	*Nitzschia clausii* Hantzsch	NITZ	This taxon is found in electrolyte-rich inlands waters and brackish coastal waters. Found in or associated with industrial effluents and is tolerant of strongly polluted conditions.
NCOM	*Nitzschia communis* Rabenhorst	NITZ	This taxon is found in brackish and electrolyte-rich waters. Tolerant of extreme levels of pollution.
NCPL	*Nitzschia capitellata* Hustedt	NITZ	This taxon is widespread in electrolyte-rich and brackish waters. Tolerant of extremely polluted conditions.
NCPR	*Navicula capitatoradiata* Germain	NAVI	This taxon is found in eutrophic fresh waters with high electrolyte content as well as brackish conditions. Tolerant to critical levels of pollution.
NCPU	*Navicymbula pusilla* Krammer	NVCB	This taxon occurs in oligo- to eutrophic waters with moderate to high electrolyte content. Common in waters with high calcium and chloride concentration.
NCRY	*Navicula cryptocephala* Kützing	NAVI	This taxon occurs in weakly acidic, oligotrophic, electrolyte poor waters and

			also in weakly alkaline, eutrophic, moderately electrolyte-rich waters. Tolerant to critical levels of pollution.
NCTE	*Navicula cryptotenella* Lange-Bertalot	NAVI	This taxon occurs in all freshwater biotopes which range between oligo- to eutrophic, with the exception of those with very high or very low electrolyte content. Tolerant only of moderately polluted conditions.
NCTO	*Navicula cryptotenelloides* Lange-Bertalot	NAVI	This taxon is found in meso- to eutrophic calcareous waters.
NDES	*Nitzschia desertorum* Hustedt	NITZ	This taxon is found in electrolyte-rich an brackish inland waters.
NDIS	*Nitzschia dissipatea* (Kützing) Grunow	NIDI	This taxon is found in waters with moderate to high electrolyte content, not present in waters with low electrolyte content.
NDME	*Nitzschia dissipata* var. *media* (Hantzsch) Grunow	NIDI	Hard water species - calcium/magnesium.
NDOI	*Navicula dutoitana* (Cholnoky)	NAVI	Rare species that inhabits acidic clean water
NDRA	*Nitzschia draveillensis* Coste & Ricard	NITZ	This taxon occurs in the plankton of eutrophic waters.
NEAF	*Neidium affine* (Ehrenberg) Pfitzer	NEID	This taxon is found in clean waters with moderate electrolyte content
NEID	*Neidium* Pfitzer	NEID	
NELE	*Nitzschia elegantula* Grunow	NITZ	This taxon is found in electrolyte-rich waters.
NEPR	*Neidium productum* (W.M.Smith) Cleve	NEID	This taxon is found is dystrophic, electrolyte poor waters
NERI	*Navicula erifuga* Lange-Bertalot	NAVI	This taxon commonly occurs in eutrophic electrolyte-rich waters, extending into brackish conditions. Tolerant to critical levels of pollution.
NETO	*Nitzschia etoshensis* Cholnoky	NITZ	This taxon occurs in electrolyte-rich and saline waters.
NFIL	*Nitzschia filiformis* (W.M.Smith) Van Heurck	NITZ	This taxon is found in waters of moderate to high electrolyte content, aslo extending into brackish conditions. Tolerant of strongly polluted conditions but not of critical levels of pollution. Occurs often in mucilage tubes.
NFON	*Nitzschia fonticola* (Grunow) Grunow	NITZ	This taxon occurs in waters with moderate to high electrolyte content. Tolerant of slightly or moderately polluted conditions. A good indicator of clean water.

NGAD	*Navigiolum adamantiforme* (Archibald) Taylor & Lange-Bertalot	NAVG	Endemic to South Africa. Occurs in alkaline, eutrophic waters with elevated electrolyte content
NGER	*Navicula germainii* Wallace	NAVI	This taxon is found in eutrophic waters, tolerant to critical levels of pollution.
NGRE	*Navicula gregaria* Donkin	NAVI	This taxon is very common in eutrophic to hypereutrophic fresh waters with moderate to high electrolyte content. Also found in brackish waters. Tolerant of strongly polluted conditions. A good indicator species for these conditions.
NHAN	*Nitzschia hantzschiana* Rabenhorst	NITZ	This taxon is found in acidic, electrolyte-poor, cool, clean waters. Common in the montane biotopes of the Drakensberg.
NHEU	*Nitzschia heufleriana* Grunow	NITZ	This taxon occurs in waters with moderate to slightly elevated electrolyte content. Not tolerant of more than moderately polluted conditions.
NHMD	*Navicula heimansioides* Lange-Bertalot	NAVI	This taxon occurs in weakly acidic to circumneutral, oligotrophic, electrolyte-poor waters.
NIAR	*Nitzschia archibaldii* Lange-Bertalot	NITZ	This taxon is common in circumneutral, slightly to moderately polluted waters with moderate electrolyte content. Reported to be tolerant of dissolved lead and zinc.
NIDI	Nitzschiae dissipatae (Section)	NIDI	
NIFR	*Nitzschia frustulum* (Kützing) Grunow	NITZ	This taxon is found in electrolyte-rich and brackish waters. Tolerant of osmotic pressure and critical levels of pollution.
NIGR	*Nitzschia gracilis* Hantzsch	NITZ	This taxon occurs in eutrophic, electrolyte-rich waters. Not tolerating more than moderately polluted conditions.
NINT	*Nitzschia intermedia* Hantzsch	NITZ	This taxon is found in the littoral zone of eutrophic rivers and lakes with moderate to high electrolyte content. Not tolerant of more than critical levels of pollution.
NIPR	*Nitzschia pura* Hustedt	NITZ	This taxon is found in weakly to moderately polluted waters with moderate electrolyte content.
NIPU	*Nitzschia pusilla* (Kützing) Grunow	NITZ	This taxon is found in a variety of eutrophic waters as well as on damp soil. Not tolerant to pollution.
NIRM	*Nitzschia irremissa* Cholnoky	NITZ	Little is know of the ecology of this taxon. Thought to be tolerant of elevated levels of pollution.
NITZ	*Nitzschia* Hassall	NITZ	

NKUZ	*Nitzschia kurzii* Rabenhorst	NITZ	This taxon occurs in brackish water, as well as mine polluted water.
NLBT	*Nitzschia liebetruthii* Rabenhorst	NITZ	This taxon occurs in electrolyte-rich and brackish waters.
NLGC	*Navicula longicephala* Hustedt	NAVI	This taxon is found in eutrophic waters with high to very high electrolyte content. Tolerant to critical levels of pollution.
NLIB	*Navicula libonensis* Schoeman	NAVI	This taxon is found in eutrophic, electrolyte-rich waters. Tolerant to critical and heavy levels of pollution.
NLIN	*Nitzschia linearis* (Agardh) W.M. Smith	NITZ	This taxon has a wide ecological range butmore common in circumneutral, oxygen rich waters of moderate to high electrolyte content. Tolerant of moderately polluted conditions.
NLSU	*Nitzschia linearis* var. *subtilis* (Grunow) Hustedt	NITZ	This taxon has a wide ecological range butmore common in circumneutral, oxygen rich waters of moderate to high electrolyte content. Tolerant of moderately polluted conditions.
NLTL	*Nitzschia lancettula* O.Müller	NITZ	This tropical/sub-tropical taxon occurs in alkaline waters.
NLTT	*Nitzschia littorea* Grunow	NITZ	This taxon occurs in waters with high salinity as well as in mine effluent.
NMCA	*Navicula microcari* Lange-Bertalot	NAVI	This taxon is common in waters with moderate to high electrolyte content.
NMCB	*Navicula microrhombus* (Cholnoky) Schoeman & Archibald	NAVI	Ecological amplitude not precisely known. Thought to occur in polluted, alkaline waters with elevated electrolyte content
NMIC	*Nitzschia microcephala* Grunow	NITZ	This taxon is found in electrolyte-rich waters with critical levels of pollution. Tolerant of changes in osmotic pressure.
NNAN	*Nitzschia nana* Grunow	NITZ	This taxon occurs in moderately polluted, electrolyte-rich and brackish waters. Tolerant of changes in osmotic pressure.
NNOT	*Navicula notha* Wallace	NAVI	This taxon is found in acidic or circumneutral, oligotrophic, electrolyte-poor waters.
NPAE	*Nitzschia paleacea* (Grunow) Grunow	NITZ	This taxon occurs in eutrophic waters with moderate to high electrolyte content. Tolerant of very heavy levels of pollution.
NPAL	*Nitzschia palea* (Kützing) W. Smith	NITZ	This commonly occurring taxon is found in eutrophic waters with moderate to high electrolyte content. Tolerant of very heavily to extremely polluted conditions.
NPML	*Nitzschia pumila* Hustedt	NITZ	Little is known of the ecology of this taxon. Found in alkaline lakes.
NPRP	*Nitzschia perspicua* Cholnoky	NITZ	This taxon is common in saline waters.

NRAD	*Navicula radiosa* Kützing	NAVI	This taxon is found in a wide variety of waters ranging from humic, weakly acidic, oligotrophic, electrolyte poor waters to strongly alkaline, eutrophic, calcareous waters. This species is however, very sensitive to organic pollution.
NRAN	*Navicula ranomafanensis* (Manguin) Metzeltin & Lange-Bertalot	NAVI	This taxon is found in acidic, oligotrophic, clean waters.
NRCH	*Navicula reichardtiana* Lange-Bertalot	NAVI	This taxon is commonly found in eutrophic waters with moderate electrolyte content as well as in calcareous waters. Tolerant to critical levels of pollution. Serves as a good indicator of these conditions.
NRCS	*Navicula recens* (Lange-Bertalot) Lange-Bertalot	NAVI	This taxon is tolerant to critical levels of pollution and is found in large eutrophic rivers with high electrolyte content, extending into brackish conditions. Often found free living and in mucilage tubes.
NREC	*Nitzschia recta* Hantzsch	NITZ	This taxon occurs in a range of waters. Not tolerant of more than moderately polluted conditions.
NREV	*Nitzschia reversa* W. Smith	NITZ	This taxon occurs in coastal and saline inland waters. Adapted to high sediment environments.
NRHY	*Navicula rhynchocephala* Kützing	NAVI	This taxon is found in oligo- to eutrophic freshwater with low to moderate electrolyte content. Occurs in greater abundance in clean waters, however, is tolerant to critical levels of pollution.
NRIE	*Navicula riediana* Lange-Bertalot & Rumrich	NAVI	This taxon is common in South Africa. Found in alkaline, eutrophic, electrolyte-rich waters.
NROS	*Navicula rostellata* Kützing	NAVI	This taxon is commonly found in eutrophic waters. Tolerant to critical levels of pollution.
NSBL	*Nitzschia sublinearis* Hustedt	NITZ	This taxon is found in slightly to moderately polluted, electrolyte-rich waters.
NSDE	*Nitzschia sinuata* var. *delognei* (Grunow) Lange-Bertalot	NITZ	This taxon is found in alkaline, meso- to eutrophic waters with moderate to high electrolyte content. Tolerant of moderately polluted conditions.
NSHR	*Navicula schroeteri* Meister	NAVI	This taxon is commonly found in eutrophic, electrolyte-rich waters. Tolerant to strongly polluted conditions.
NSIG	*Nitzschia sigma* (Kützing) W.M. Smith	NITZ	This taxon is found in eutrophic, electrolyte-rich waters and extending into brackish and coastal waters.

NSIT	*Nitzschia sinuata* var. *tabellaria* Grunow	NITZ	This taxon is found in circumneutral, mesotrophic waters of moderate electrolyte content. Tolerant of critical levels of pollution.
NSRH	*Navicula subrhynchocephala* Hustedt	NAVI	This taxon commonly occurs in electrolyte- rich waters in warmer climatic regions.
NSSY	*Navicula schroeteri* var. *symmetrica* (Patrick) Lange-Bertalot	NAVI	This taxon is commonly found in eutrophic, electrolyte-rich waters. Tolerant to strongly polluted conditions.
NSYM	*Navicula symmetrica* Patrick	NAVI	This taxon is commonly found in eutrophic, electrolyte-rich waters. Tolerant to strongly polluted conditions.
NTEN	*Navicula tenelloides* Hustedt	NAVI	This aerophytic taxon is found in waters with a wide range of electrolyte content and varied trophic status. Tolerant to extremely polluted conditions.
NTPT	*Navicula tripunctata* (O.F.Müller) Bory	NAVI	This taxon is tolerant to critical levels of pollution and is a good indicator for eutrophic conditions with moderate to high electrolyte content. Often found free living and in mucilage tubes.
NTRV	*Navicula trivialis* Lange-Bertalot	NAVI	This epipelic taxon occurs in eutrophic waters with moderate electrolyte content. Tolerant of desiccation and strongly polluted conditions.
NUMB	*Nitzschia umbonata* (Ehrenberg) Lange-Bertalot	NITZ	This taxon is common in eutrophic electrolyte-rich waters. Tolerant of extreme levels of pollution.
NVCB	*Navicymbula* Krammer	NVCB	
NVDA	*Navicula vandamii* Schoeman & Archibald	NAVI	This taxon is found in alkaline, eutrophic waters with high electrolyte content.
NVEN	*Navicula veneta* Kützing	NAVI	This taxon is found in heavily eutrophic, electrolyte-rich to brackish waters. Highly tolerant to pollution, often dominant in industrially impacted water.
NVIR	*Navicula viridula* (Kützing) Ehrenberg	NAVI	This taxon is infrequently found. Occurs in eutrophic waters. Tolerant to critical levels of pollution. An epilithic and epipelic species also occurring on detritus and submerged macrophytes.
NVLC	*Nitzschia valdecostata* Lange-Bertalot & Simonsen	NITZ	This taxon occurs in electrolyte-rich waters, favouring high concentrations of sulphates and carbonates.
NZAN	*Navicula zanoni* Hustedt	NAVI	This tropical to sub-tropical taxon is found in alkaline waters of South Africa.
NZCL	*Nitzschia closterium* (Ehrenberg) W. Smith	NITZ	Thin silica frustule, destroyed during acid cleaning, sample preparation, Indicate extreme salinity.

NZRA	*Nitzschia radicula* Hustedt	NITZ	This taxon is found in slightly to moderately polluted, electrolyte-rich waters.
NZSU	*Nitzschia supralitorea* Lange-Bertalot	NITZ	This taxon is found in eutrophic waters with moderate to moderately high electrolyte content. Also occurring in supralittoral sites, tolerant of osmotic fluctuations and strongly polluted conditions.
PACR	*Pinnularia acrospheria* W. Smith	PINU	This epipelic taxon is found in circumneutral waters with moderate electrolyte content. Occurring mostly in the tropics.
PCLT	*Placoneis clementis* (Grunow) Cox	PLAC	Found in electrolyte rich, brackish waters.
PDIC	*Placoneis dicephala* (W. Smith) Mereschkowsky	PLAC	This taxon is found on the sediments in a range of waters and is tolerant of moderate levels of pollution
PDIV	*Pinnularia divergens* W.M.Smith	PINU	This taxon is found in mountainous areas. Occurs in acidic, oligotrophic waters with low electrolyte content.
PDUN	*Pinnularia divergens* var. *undulata* (Peragello & Heribaud) Hustedt	PINU	This taxon has an isolated distribution, occurring in acidic, oligotrophic, electrolyte-poor waters.
PELG	*Placoneis elginensis* (Gregory) Cox	PLAC	This taxon occurs in unpolluted or slightly polluted dystrophic waters
PELO	*Pleurosigma elongatum* W.Smith	PLSG	Cosmopolitan taxon occurring in brackish inland waters
PGIB	*Pinnularia gibba* Ehrenberg	PINU	This taxon is common is springs and small streams. Occurs in waters with low to moderate electrolyte content.
PINU	*Pinnularia* Ehrenberg	PINU	
PJOC	*Pinnularia joculata* (Manguin) Krammer	PINU	The ecology of this tropical taxon is unknown, possibly occurring in acidic waters.
PLAC	*Placoneis* Mereschkowsky	PLAC	
PLEN	*Planothidium engelbrechtii* (Cholnoky) Round & Bukhtiyarova	PLTD	This taxon is abundant in inland saline waters with high electrolyte content. Tolerant to critical levels of organic pollution
PLFR	*Planothidium frequentissimum* (Lange-Bertalot) Lange-Bertalot	PLTD	This taxon is found in circumneutral to alkaline water with moderate to high electrolyte content. Tolerant of critical levels of pollution
PLHU	*Platessa hustedtii* (Krasske) Lange-Bertalot	PTSA	This taxon is more often found in small, circumneutral, oligotrophic streams with low electrolyte content.
PLSG	*Pleurosigma* W. Smith	PLSG	
PLTD	*Planothidium* Round & Bukhtiyarova	PLTD	

PMRO	*Pinnularia microstauron* (Ehrenberg) Cleve var. *rostrata* Krammer	PINU	This taxon is found in clean, circumneutral, oligotrophic waters with low electrolyte content.
POBG	*Psammothidium oblongellum* (Oestrup) Van de Vijver (*Platessa oblongella*)	PSMT	Check data, analyse new values and compare
PPLC	*Placoneis placentula* (Ehrenberg) Heinzerling	PLAC	This taxon occurs in unpolluted or slightly polluted water with moderate to high electrolyte content
PRST	*Planothidium rostratum* (Østrup) Lange-Bertalot	PLTD	This taxon is found in circumneutral to alkaline water with moderate to low electrolyte content. More often attached to plants than stones
PSAL	*Pleurosigma salinarum* (Grunow) Cleve & Grunow	PLSG	Cosmopolitan taxon occurring in brackish and saline inland waters
PSBR	*Pseudostaurosira brevistriata* (Grunow) D.M. Williams & Round	PSST	Brackish species. Eutrophic. Cluster of chain forming colony
PSBV	*Pinnularia subbrevistriata* Krammer	PINU	This tropical to sub-tropical taxon occurs in moderately polluted conditions.
PSCA	*Pinnularia subcapitata* Gregory	PINU	This taxon is found in oligotrophic, electrolyte-poor waters.
PSMT	*Psammothidium* Bukhtiyarova & Round	PSMT	
PSST	*Pseudostaurosira* (Grunow) D.M. Williams & Round	PSST	
PTLA	*Planothidium lanceolatum* (Brébisson) Lange-Bertalot	PLTD	This taxon is tolerant to elevated nutrient levels but not tolerant to more than moderate levels of pollution.
PTSA	*Platessa* Lange Bertalot	PTSA	
PVIF	*Pinnularia viridiformis* Krammer	PINU	This taxon is one of the most commonly occurring in its genus. Found in oligo- to mesotrophic waters with low to moderate electrolyte content.
PVIR	*Pinnularia viridis* (Nitzsch) Ehrenberg	PINU	This taxon is found in circumneutral, oligo- to mesotrophic waters with low to moderate electrolyte content.
RABB	*Rhoicosphenia abbreviata* (C. Agardh) Lange-Bertalot	RHOI	This taxon is found din electrolyte-rich as well as brackish inland waters. Tolerant to critical levels of pollution. Attached to substratum by means of antapically attached mucilage stalks.
REIM	*Reimeria* Kociolek & Stoermer	REIM	
RGBL	*Rhopalodia gibberula* (Ehrenberg) O.Müller	RHOP	This taxon is found in waters with moderate to high electrolyte content. Tolerant of elevated water temperatures.
RGIB	*Rhopalodia gibba* (Ehrenberg) O. Müller	RHOP	This taxon occurs in standing and slow flowing waters, especially springs, with moderate to high electrolyte content.
RHOI	Rhoicosphenia Grunow	RHOI	
RHOP	Rhopalodia Müller	RHOP	

RMUS	*Rhopalodia musculus* (Kützing) O. Müller	RHOP	This taxon is found in electrolyte-rich and brackish waters.
ROPE	*Rhopalodia operculata* (Agardh) Hakansson	RHOP	This taxon occurs in waters of moderate to high electrolyte content. Also found in thermal mineral springs.
RSIN	*Reimeria sinuata* (Gregory) Kociolek & Stoermer	REIM	This aerophilic taxon is found in montane areas, occurring in springs, streams and on bryophytes.
RUNI	*Reimeria uniseriata* Sala, Guerrero & Ferrario	REIM	This taxon is found in alkaline, meso- to eutrophic waters with moderate electrolyte content. Able to grow under light limiting conditions.
SAGA	*Stephanodiscus agassizensis* Håkansson & Kling	STEP	Taxon found in turbid eutrophic rivers and lakes with high electrolyte content
SANC	*Stenopterobia anceps* (Lewis) Brébisson	STEN	This taxon occurs in oligotrophic, acidic waters with moderate electrolyte content.
SANG	*Surirella angusta* Kützing	SURI	This taxon is found in eutrophic waters with moderate electrolyte content. Highly tolerant to dissolved metals, often found in urban and mining polluted effluent.
SBRE	*Surirella brebissonii* Krammer & Lange-Bertalot	SURI	This taxon is found in waters of moderate to high electrolyte content also extending into brackish conditions.
SCON	*Staurosira construens* Ehrenberg	STRS	Chain-forming planktonic species occurring in good quality waters
SCRU	*Surirella crumena* Brébisson	SURI	This taxon is found in electrolyte-rich and brackish waters. More common in coastal waters.
SELI	*Staurosira elliptica* (Schumann) Williams & Round	STRS	Chain-forming planktonic species. High electrolyte conditions.
SELL	*Sellaphora* Mereschkowsky	SELL	
SHAN	*Stephanodiscus hantzschii* Grunow	STEP	Planktonic taxon found in waters with elevated electrolyte content
SIDE	*Simonsenia delognei* Lange-Bertalot	SIMO	This taxon is found in electrolyte-rich and brackish waters. Tolerant to osmotic fluctuations. Also found in soil.
SIMO	*Simonsenia* Lange-Bertalot	SIMO	
SMNA	*Seminavis* D.G. Mann	SMNA	
SMST	*Seminavis strigosa* (Hustedt) Danieledis & Economou-Amilli	SMNA	This taxon is found abundant in saline waters.
SOVI	*Surirella ovalis* Brébisson	SURI	This taxon is found in electrolyte-rich environments.
SPHO	*Stauroneis phoenicenteron* (Nitzsch) Ehrenberg	STAU	This taxon is found in eutrophic waters, sometimes polluted water.
SPIN	*Staurosirella pinnata* (Ehrenberg) Williams & Round	STRL	Chain-forming species. Clean waters with moderate to high electrolyte content
SPUP	*Sellaphora pupula* (Kützing) Mereschkowksy	SELL	This cosmopolitan taxon occurs in a broad range of waters. Indicator of high nutrients and organic waste

SSEM	*Sellaphora seminulum* (Grunow) D.G. Mann	SELL	This cosmopolitan taxon occurs in a braod range of waters. Indicator of high nutrients and organic waste
SSMI	*Stauroneis smithii* Grunow	STAU	The ecology of this taxon is uncertain. Only isolated examples of this taxon is found in samples. It is reported from eutrophic waters with moderate electrolyte water, and also from waters with low electrolyte content.
SSTM	*Sellaphora stroemii* (Hustedt) D.G. Mann	SELL	This taxon is found in cold electrolyte rich waters of streams, springs and waterfalls
STAN	Stauroneis anceps Ehrenberg	STAU	This taxon occurs in oligo- to mesotrophic waters with low to moderate electrolyte content.
STAU	*Stauroneis* Ehrenberg	STAU	
STDE	*Stenopterobia delicatissima* (Lewis) Brébisson	STEN	This taxon occurs in oligotrophic, acidic waters with moderate electrolyte content.
STEN	*Stenopterobia* Brébisson	STEN	
STEP	*Stephanodiscus* Ehrenberg	STEP	
STMI	*Stephanodiscus minutulus* (Kützing) Cleve & Moller	STEP	Planktonic taxon found in strongly polluted waters with high electrolyte content
STRL	Staurosirella D.M. Williams & F.E. Round emend Morales	STRL	
STRS	*Staurosira* (C.G. Ehrenberg) D.M. Williams &. Round	STRS	
SURI	*Surirella* Turpin	SURI	
TABE	*Tabellaria* Ehrenberg	TABE	
TABU	*Tabularia* D.M. Williams & Round	TABU	
TAPI	*Tryblionella apiculata* Gregory	TRYB	This taxon is tolerant of strongly polluted conditions. Found in electrolyte-rich waters.
TCAL	*Tryblionella calida* (Grunow) D.G. Mann	TRYB	This taxon occurs in eutrophic waters with elevated electrolyte content. Favours standing waters.
TCOA	*Tryblionella coarctata* (Grunow) D.G. Mann	TRYB	This taxon occurs in brackish waters.
TDEB	*Tryblionella debilis* Arnott	TRYB	This taxon favours biotopes subject to osmotic fluctuation, including mosses, rock crevices and damp soil. Rarely abundant.
TFAS	*Tabularia fasciculata* (Agardh) D.M. Williams & Round	TABU	This taxon commonly occurs in brackish waters (estuaries) and is often found in critically polluted industrial wastewater
TFLO	*Tabellaria flocculosa* (Roth) Kützing	TABE	Taxon found in oligotrophic acidic waters, common in the Magaliesberg region of the Western cape.
TGRL	*Tryblionella gracilis* W. Smith	TRYB	This taxon occurs in brackish waters, sometimes in electrolyte rich waters. Tolerant of osmotic fluctuations.

THAL	*Thalassiosira* P.T. Cleve	THAL	
THUN	*Tryblionella hungarica* (Grunow) D.G. Mann	TRYB	This taxon is tolerant of strongly polluted conditions. Found in electrolyte-rich waters, extending into brackish conditions.
TLEV	*Tryblionella levidensis* W.M. Smith	TRYB	This subtropical taxon occurs in waters ranging from moderate electrolyte content to brackish conditions. Tolerant of strongly polluted conditions.
TLIT	*Tryblionella littoralis* (Grunow) D.G. Mann	TRYB	This taxon occurs in tidal zones and biotopes influenced by brackish waters.
TPSN	*Thalassiosira pseudonana* Hasle & Heimdal	THAL	Planktonic taxon found in waters with elevated electrolyte content, also occurring in intertidal zones of marine waters
TRYB	*Tryblionella* W. Smith	TRYB	
TWEI	*Thalassiosira weissflogii* (Grunow) Fryxell & Hasle	THAL	Planktonic taxon found in waters with elevated electrolyte content
UACU	*Ulnaria acus* (Kützing) Aboal	ULNA	Adapted to high flow conditions
UAMP	*Ulnaria amphirhynchus* (Ehrenberg) Compère & Bukhtiyarova	ULNA	Adapted to high flow conditions
ULNA	*Ulnaria* Compère	ULNA	
UULN	*Ulnaria ulna* (Nitzsch) Compère	ULNA	Adapted to high flow conditions, found in meso- eutrophic alkaline waters
UUNG	*Ulnaria ungeriana* (Grunow) Compère	ULNA	Found in sub-tropical weakly alkaline waters with lower organic material

Table A.3: Trophic preference and tolerance classes for species for EC, DIN and PO_4^{3-}, as well as combined trophic preference and tolerance classes.

CODE	Salinity (EC) preference	Salinity (EC) Tolerance class	PO_4^{3-} trophic preference	PO_4^{3-} tolerance class	DIN trophic preference	DIN tolerance class	Combined trophic preference	Combined tolerance class
ADEG	mesohaline	generalist	eutrophic	generalist	mesotrophic	generalist	meso-eutrophic	Generalist
ADMI	oligohaline	generalist	mesotrophic	intermediate	oligotrophic	specialist	oligo-mesotrophic	Intermediate
ADRI	oligohaline	specialist	meso-eutrophic	generalist	oligotrophic	specialist	oligo-mesotrophic	Specialist
ADSA	oligo-mesohaline	intermediate	mesotrophic	generalist	oligotrophic	specialist	oligo-mesotrophic	Intermediate
AEXI	mesohaline	specialist	eutrophic	generalist	mesotrophic	intermediate	meso-eutrophic	Intermediate
AMMO	mesohaline	generalist	mesotrophic	generalist	oligotrophic	specialist	mesotrophic	Generalist
APED	mesohaline	generalist	mesotrophic	generalist	oligotrophic	intermediate	mesotrophic	Generalist
AUGA	oligo-mesohaline	generalist	mesotrophic	generalist	oligotrophic	intermediate	oligo-mesotrophic	Generalist
AUGR	meso-euryhaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
AVEN	euryhaline	generalist	eutrophic	generalist	meso-eutrophic	generalist	eutrophic	Generalist
CAFF	oligo-mesohaline	generalist	oligo-mesotrophic	intermediate	oligotrophic	specialist	oligo-mesotrophic	Intermediate
CINV	mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist

CKPP	oligo-mesohaline	intermediate	mesotrophic	intermediate	oligotrophic	specialist	oligo-mesotrophic	Intermediate
CMED	meso-euryhaline	specialist	meso-eutrophic	generalist	oligotrophic	specialist	meso-eutrophic	Specialist
CMEN	meso-euryhaline	generalist	eutrophic	generalist	oligo-mesotrophic	generalist	meso-eutrophic	Generalist
CMLF	oligo-mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	specialist	mesotrophic	Generalist
CPED	mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
CPLA	mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
CPLE	mesohaline	intermediate	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Intermediate
CTGL	oligo-mesohaline	specialist	mesotrophic	generalist	oligotrophic	specialist	oligo-mesotrophic	Specialist
CTUM	oligo-mesohaline	intermediate	mesotrophic	intermediate	oligotrophic	specialist	oligo-mesotrophic	Intermediate
DCOF	mesohaline	intermediate	meso-eutrophic	generalist	mesotrophic	intermediate	meso-eutrophic	Intermediate
DCOF	oligohaline	intermediate	mesotrophic	generalist	mesotrophic	intermediate	mesotrophic	Intermediate
DVUL	mesohaline	generalist	mesotrophic	generalist	oligotrophic	intermediate	mesotrophic	Generalist
EARC	meso-euryhaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
EMIN	mesohaline	specialist	mesotrophic	specialist	oligotrophic	intermediate	mesotrophic	Generalist
ENCM	mesohaline	generalist	mesotrophic	specialist	oligotrophic	specialist	mesotrophic	Specialist
ENLS	oligo-mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	mesotrophic	Intermediate
ENMI	oligo-mesohaline	generalist	mesotrophic	generalist	oligotrophic	intermediate	oligo-mesotrophic	Generalist
EOLI	mesohaline	generalist	eutrophic	generalist	oligo-mesotrophic	generalist	meso-eutrophic	Generalist
EOMI	mesohaline	generalist	mesotrophic	generalist	oligo-mesotrophic	intermediate	mesotrophic	Generalist
ESBM	mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	generalist	meso-eutrophic	Generalist
ESOR	mesohaline	specialist	mesotrophic	intermediate	oligotrophic	specialist	mesotrophic	Specialist
FBCP	oligo-mesohaline	intermediate	mesotrophic	generalist	oligotrophic	intermediate	oligo-mesotrophic	Intermediate
FBID	oligohaline	intermediate	mesotrophic	intermediate	oligotrophic	specialist	oligo-mesotrophic	Intermediate
FBRE	mesohaline	generalist	eutrophic	generalist	mesotrophic	intermediate	meso-eutrophic	Generalist
FCAP	oligo-mesohaline	specialist	mesotrophic	specialist	oligotrophic	intermediate	oligo-mesotrophic	Specialist
FCVA	oligo-mesohaline	intermediate	mesotrophic	generalist	oligotrophic	specialist	oligo-mesotrophic	Intermediate
FSAP	mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
FULN	oligo-mesohaline	generalist	mesotrophic	generalist	oligotrophic	intermediate	oligo-mesotrophic	Generalist
GEXL	oligohaline	generalist	oligo-mesotrophic	intermediate	oligotrophic	specialist	meso-eutrophic	Intermediate
GLST	oligohaline	specialist	mesotrophic	intermediate	oligotrophic	specialist	oligo-mesotrophic	Intermediate
GMIN	oligo-mesohaline	generalist	mesotrophic	intermediate	oligotrophic	specialist	oligo-mesotrophic	Specialist
GOMP sp	mesohaline	intermediate	eutrophic	generalist	oligo-mesotrophic	intermediate	oligo-mesotrophic	Intermediate
GPAR	oligo-mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	mesotrophic	Generalist
GPSA	euryhaline	generalist	eutrophic	generalist	eutrophic	generalist	eutrophic	Generalist
GPUM	oligo-mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	mesotrophic	Generalist

GVNU	oligo-mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	mesotrophic	Generalist
MAPE	mesohaline	generalist	eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
MVAR	mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
NACI	mesohaline	intermediate	mesotrophic	intermediate	oligotrophic	specialist	mesotrophic	Intermediate
NADF	mesohaline	intermediate	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Intermediate
NAMP	mesohaline	generalist	eutrophic	generalist	meso-eutrophic	generalist	eutrophic	Generalist
NANT	mesohaline	intermediate	eutrophic	generalist	meso-eutrophic	generalist	eutrophic	Generalist
NARV	mesohaline	generalist	eutrophic	generalist	mesotrophic	generalist	meso-eutrophic	Generalist
NAVS	mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
NCPL	meso-euryhaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	generalist	meso-eutrophic	Generalist
NCPR	oligo-mesohaline	intermediate	mesotrophic	generalist	oligotrophic	specialist	oligo-mesotrophic	Intermediate
NCRY	oligo-mesohaline	generalist	mesotrophic	generalist	oligotrophic	intermediate	oligo-mesotrophic	Generalist
NCTE	mesohaline	generalist	mesotrophic	generalist	oligotrophic	intermediate	mesotrophic	Generalist
NDIS	oligo-mesohaline	generalist	mesotrophic	intermediate	oligotrophic	specialist	oligo-mesotrophic	Intermediate
NERI	mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
NFIL	meso-euryhaline	generalist	eutrophic	generalist	oligotrophic	generalist	meso-eutrophic	Generalist
NFON	meso-euryhaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
NGER	oligo-mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	mesotrophic	Generalist
NGRE	meso-euryhaline	generalist	eutrophic	generalist	mesotrophic	intermediate	eutrophic	Generalist
NIAR	meso-euryhaline	generalist	mesotrophic	intermediate	oligotrophic	intermediate	meso-eutrophic	Intermediate
NIFR	mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
NIGR	meso-euryhaline	specialist	mesotrophic	intermediate	oligotrophic	specialist	meso-eutrophic	Specialist
NINC	meso-euryhaline	generalist	eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
NIPU	mesohaline	intermediate	mesotrophic	intermediate	oligotrophic	specialist	mesotrophic	Intermediate
NLBT	mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
NLIN	mesohaline	generalist	mesotrophic	intermediate	oligotrophic	generalist	mesotrophic	Generalist
NLSU	oligo-mesohaline	intermediate	mesotrophic	intermediate	oligotrophic	specialist	oligo-mesotrophic	Intermediate
NMEN	mesohaline	generalist	mesotrophic	generalist	oligotrophic	intermediate	mesotrophic	Generalist
NMIN	mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	specialist	meso-eutrophic	Generalist
NPAE	mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	generalist	meso-eutrophic	Generalist
NPAL	mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	generalist	meso-eutrophic	Generalist
NPHO	oligo-mesohaline	intermediate	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	mesotrophic	Intermediate
NRCS	meso-euryhaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
NROS	mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist

NSHR	mesohaline	generalist	meso-eutrophic	generalist	mesotrophic	intermediate	meso-eutrophic	Generalist
NSLT	mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
NSYM	mesohaline	generalist	eutrophic	generalist	oligo-mesotrophic	generalist	meso-eutrophic	Generalist
NTEN	mesohaline	generalist	mesotrophic	intermediate	oligo-mesotrophic	intermediate	mesotrophic	Intermediate
NVDA	oligo-mesohaline	intermediate	mesotrophic	specialist	oligotrophic	specialist	oligo-mesotrophic	Specialist
NVEN	meso-euryhaline	generalist	eutrophic	generalist	oligo-mesotrophic	generalist	meso-eutrophic	Generalist
NZSU	mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
PDAU	mesohaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
PENG	mesohaline	intermediate	mesotrophic	generalist	oligo-mesotrophic	intermediate	mesotrophic	Intermediate
PLFR	meso-euryhaline	generalist	meso-eutrophic	generalist	mesotrophic	intermediate	meso-eutrophic	Generalist
POBG	oligo-mesohaline	specialist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	mesotrophic	Intermediate
PTRO	oligo-mesohaline	generalist	mesotrophic	generalist	oligo-mesotrophic	intermediate	mesotrophic	Generalist
RABB	mesohaline	generalist	eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
RUNI	oligo-mesohaline	intermediate	mesotrophic	generalist	oligotrophic	specialist	oligo-mesotrophic	Intermediate
SAGA	meso-euryhaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
SBRE	meso-euryhaline	generalist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Generalist
SELI	meso-euryhaline	generalist	meso-eutrophic	generalist	mesotrophic	specialist	meso-eutrophic	Generalist
SHAN	meso-euryhaline	specialist	meso-eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Intermediate
SMST	mesohaline	intermediate	eutrophic	generalist	oligo-mesotrophic	intermediate	meso-eutrophic	Intermediate
SPAV	mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
SPUP	mesohaline	generalist	eutrophic	generalist	meso-eutrophic	generalist	eutrophic	Generalist
SSEM	mesohaline	generalist	eutrophic	generalist	meso-eutrophic	generalist	eutrophic	Generalist
TAPI	meso-euryhaline	generalist	meso-eutrophic	generalist	oligotrophic	intermediate	meso-eutrophic	Generalist
TPSN	mesohaline	generalist	meso-eutrophic	generalist	oligotrophic	specialist	meso-eutrophic	Generalist

Tabel A.4: Index scores of the final selected indices across all sites, accompanied by the IPS scores and measured water quality for the relevant parameters used to create the indices (EC, DIN and PO_4^{3-}).

Site	IPS	**WAI-AVG (present study)**	KBI-A (present study)	KBI-N (present study)	KBI-O (present study)	KBI-I (present study)	WAI-EC (present study)	WAI-DIN (present study)	WAI-OP (present study)	EC	DIN	PO_4^{3-}
D1-1	9.89	**11.77**	**10.85**	**7.08**	**9.34**	**13.50**	11.60	19.28	6.60	229	0.141	0.062
D1-2	14.79	**12.88**	**11.98**	**7.89**	**10.15**	**14.93**	12.98	19.90	7.75	315	0.083	0.022
D1-3	12.92	**13.56**	**12.30**	**6.69**	**10.13**	**16.19**	13.54	19.86	9.08	332	0.192	0.018
D1-4	10.38	**14.19**	**11.03**	**5.23**	**9.27**	**14.81**	14.40	19.80	9.94	480	0.188	0.024
D1-5	12.19	**14.37**	**12.38**	**8.16**	**10.39**	**15.48**	14.65	19.80	10.18	510	0.113	0.055
D1-6	15.75	**14.46**	**13.05**	**5.97**	**10.28**	**17.97**	14.83	19.95	10.12	533	0.601	0.064
D1-7	9.22	**12.25**	**9.41**	**1.96**	**9.27**	**13.20**	11.92	19.29	7.60	214	0.398	0.085

D1-8	9.66	12.82	8.88	3.80	8.47	11.63	13.36	19.10	7.72	147	0.203	0.063
D1-9	11.28	12.20	10.57	6.20	8.80	13.64	12.07	19.10	7.42	148	0.337	0.033
D1-10	15.25	12.76	11.79	7.37	10.10	14.86	12.81	19.71	7.75	178	0.146	0.153
D1-11	15.34	12.82	11.92	7.64	10.37	14.83	13.00	19.70	7.72	323	0.533	0.079
D1-12	12.24	13.03	10.88	5.24	9.30	14.49	13.63	18.96	8.11	323	0.533	0.079
D1-13	13.79	15.05	8.90	4.34	10.45	10.41	16.01	19.97	10.43	278	0.126	0.012
D1-14	13.83	13.76	10.31	4.97	10.16	13.05	13.76	19.92	9.38	273	0.132	0.013
D1-15	13.85	13.11	10.20	3.65	9.97	13.59	12.45	19.61	9.22	284	0.099	0.006
D1-16	13.73	13.08	9.08	2.76	8.81	12.38	12.53	19.15	9.38	291	0.06	0.024
D1-17	15.32	14.24	10.72	3.88	10.49	14.25	14.03	19.96	10.40	290	0.103	0.019
D1-18	14.29	14.60	10.90	5.71	10.56	13.68	15.17	19.96	10.12	293	0.06	0.017
D1-19	13.55	14.35	9.44	3.59	10.25	11.96	14.62	19.87	10.10	299	0.144	0.006
D1-20	13.84	13.80	10.49	5.09	10.25	13.32	13.92	19.74	9.43	300	0.148	0.03
D1-21	13.86	13.93	10.50	4.33	10.21	13.73	13.83	19.72	9.92	308	0.093	0.018
D1-22	13.26	14.34	9.71	4.59	10.22	12.02	14.80	19.80	9.92	294	0.099	0.006
D1-23	12.01	13.88	9.20	3.46	9.93	11.70	13.98	19.83	9.52	276	0.213	0.023
D1-24	6.36	10.99	7.54	2.98	7.68	9.75	10.50	17.96	6.58	249	0.19	0.05
D1-25	14.46	11.84	12.18	8.81	10.19	14.86	11.46	19.83	6.55	342	0.06	0.041
D1-26	8.13	11.80	8.72	3.70	8.41	11.38	11.06	19.03	7.49	379	0.094	0.025
D1-27	10.87	12.47	9.47	3.68	8.83	12.68	11.29	19.58	8.73	440	0.06	0.026
D1-28	11.15	12.12	9.75	4.40	9.13	12.73	11.11	19.12	8.28	454	0.06	0.032
D1-29	8.02	11.66	9.43	3.32	8.88	12.76	11.24	18.76	7.08	430	0.092	0.027
D1-30	11.95	12.55	10.91	4.95	9.54	14.59	12.08	19.31	8.25	433	0.168	0.055
D1-31	14.56	13.56	10.64	4.26	11.65	13.33	12.64	19.84	10.11	440	0.164	0.053
D1-32	13.43	12.82	11.26	5.82	11.10	14.05	12.17	19.84	8.54	515	0.145	0.043
D1-33	11.37	11.96	10.04	4.60	9.37	13.10	11.58	18.38	7.81	247	0.082	0.06
D1-34	7.31	12.19	9.30	3.60	8.66	12.47	11.96	19.04	7.55	379	0.06	0.054
D1-35	10.22	12.08	9.45	4.64	9.25	11.95	11.89	18.92	7.41	227	0.186	0.041
D1-36	8.93	13.08	11.39	6.71	9.20	14.83	13.49	19.34	8.13	233	0.134	0.016
D1-37	11.70	13.35	11.29	6.45	9.37	14.66	13.80	19.85	8.19	234	0.178	0.022
D1-38	11.37	13.81	10.72	5.20	9.92	13.89	14.45	19.88	8.74	241	0.198	0.026
D1-39	10.53	13.85	10.02	3.95	9.96	13.09	14.12	19.72	9.34	247	0.18	0.03
D1-40	12.47	13.90	11.80	6.59	10.11	15.25	14.21	19.70	9.41	239	0.215	0.043
D1-41	11.68	14.45	9.14	4.42	10.33	10.92	14.94	19.90	9.98	238	0.466	0.044
D1-42	11.35	14.13	10.63	6.26	10.05	13.10	14.69	19.92	9.34	238	0.465	0.085
D1-43	11.69	14.05	11.61	7.60	9.97	14.43	14.49	19.90	9.37	238	0.266	0.127
D1-44	14.49	13.59	12.58	6.37	10.34	16.80	13.41	19.88	9.29	231	0.283	0.038
D1-45	14.60	12.17	12.22	8.54	10.25	15.05	11.86	19.85	7.05	240	0.227	0.028
D1-46	11.36	12.94	10.67	5.49	9.59	13.80	13.03	19.64	8.04	236	0.217	0.04
D1-47	8.38	10.93	7.50	2.32	7.84	9.93	10.49	18.66	5.93	325	0.606	0.084
D1-48	11.26	10.71	9.95	4.47	9.03	13.15	10.38	19.07	5.12	590	1.237	0.122
D1-49	12.65	9.87	9.95	2.48	9.40	13.95	9.98	17.83	4.07	740	2.305	0.16
D1-50	10.34	10.06	8.50	2.16	7.79	12.03	9.02	17.80	5.73	806	3.225	0.186
D1-51	12.00	9.19	8.53	1.80	7.76	12.28	8.05	16.51	5.26	774	2.681	0.178

D1-52	12.52	7.46	7.95	1.33	7.30	11.58	7.18	13.31	3.59	840	2.937	0.199
D1-53	9.19	8.45	8.18	1.39	8.04	11.64	7.64	14.68	4.92	845	2.433	0.239
D1-54	9.04	8.88	7.68	1.80	7.66	10.63	8.73	15.08	4.62	819	2.122	0.199
D1-55	5.84	9.46	6.92	1.92	6.91	9.43	8.44	16.95	5.28	855	1.322	0.315
D1-56	13.76	10.88	11.83	8.49	9.71	14.57	9.78	19.72	5.83	815	1.32	0.185
D1-57	4.46	10.08	6.00	1.59	6.35	8.03	8.69	18.28	5.81	818	0.443	0.235
D1-58	7.55	10.22	7.02	1.62	7.06	9.70	8.53	19.31	5.64	844	0.818	0.326
D1-59	12.86	10.94	11.56	8.64	9.36	14.12	10.38	19.30	5.61	336	0.632	0.135
D1-60	11.38	10.32	9.27	4.14	7.95	12.49	8.96	18.97	5.70	543	0.324	0.08
D1-61	11.93	9.16	7.36	1.57	6.19	10.83	6.71	17.79	5.80	712	1.513	0.107
D1-62	10.47	9.46	7.78	2.84	6.82	10.74	8.05	17.13	5.60	705	2.247	0.104
D1-63	10.96	8.49	7.03	1.78	6.23	10.05	6.27	16.27	5.49	804	1.52	0.082
D1-64	11.97	8.70	6.97	1.56	6.13	10.10	6.78	16.10	5.60	861	1.977	0.21
D1-65	13.93	8.71	7.24	1.56	6.14	10.63	6.57	15.96	5.98	803	1.951	0.244
D1-66	12.97	8.60	7.26	1.71	6.25	10.54	6.48	16.07	5.68	845	0.813	0.274
D1-67	8.88	8.76	7.32	2.59	6.76	9.96	7.51	17.21	4.16	858	0.06	0.307
D1-68	10.26	9.44	7.88	2.56	6.90	11.03	7.91	18.10	5.00	793	0.073	0.073
D1-69	8.34	11.08	7.93	3.26	7.78	10.34	10.69	18.46	6.27	771	0.088	0.109
D1-70	3.62	10.51	4.76	1.10	6.07	5.94	9.79	18.46	5.64	823	0.06	0.111
D1-71	11.47	11.02	11.22	8.23	9.10	13.77	10.39	19.30	5.83	668	1.594	0.203
D1-72	12.85	9.90	8.14	2.77	6.90	11.45	8.65	17.39	5.99	548	0.06	0.078
D1-73	12.66	9.07	7.33	1.95	6.25	10.56	7.45	16.48	5.61	744	0.756	0.08
D1-74	13.07	8.64	7.22	1.61	6.31	10.49	6.53	15.85	5.92	714	1.009	0.112
D1-75	11.62	8.52	6.85	1.37	6.09	9.97	6.32	16.01	5.68	809	1.058	0.045
D1-76	8.88	8.35	5.94	1.32	5.96	8.25	6.18	16.04	5.34	902	1.126	0.117
D1-77	9.05	8.17	6.06	1.11	5.83	8.65	6.18	15.58	5.16	908	0.994	0.158
D1-78	7.55	9.27	6.98	2.28	6.58	9.54	8.31	17.17	4.71	906	0.481	0.235
D1-79	8.45	9.66	6.71	1.41	6.08	9.67	8.51	18.26	4.82	1010	0.552	0.257
D1-80	7.98	10.05	7.44	3.58	6.77	9.71	9.59	17.90	4.95	821	0.06	0.159
D1-81	7.37	9.97	7.18	2.03	6.43	10.14	7.94	19.26	5.66	905	0.087	0.104
D1-82	9.68	10.14	7.10	2.27	7.19	9.48	7.93	19.39	6.06	973	0.079	0.144
D1-83	14.77	9.79	5.30	1.45	6.32	6.71	8.05	18.94	5.26	567	1.477	0.171
D1-84	9.95	10.78	6.64	1.38	5.44	9.88	9.79	19.45	5.73	521	0.06	0.066
D1-85	10.32	10.61	7.02	1.62	6.70	9.88	9.39	19.44	5.68	555	1.745	0.281
D1-86	9.62	10.33	6.57	1.68	6.73	8.94	9.26	18.87	5.43	704	0.619	0.109
D1-87	8.06	9.21	6.24	1.71	6.22	8.50	7.97	17.78	4.51	741	0.267	0.033
D1-88	6.51	8.99	6.39	1.52	6.64	8.71	7.16	17.03	5.35	820	0.077	0.058
D1-89	11.48	8.48	7.06	1.27	6.34	10.31	6.66	15.98	5.20	916	0.529	0.074
D1-90	8.09	10.61	7.01	2.00	6.80	9.62	9.58	19.08	5.73	935	0.156	0.093
D1-91	8.56	9.90	7.10	1.62	6.61	10.09	8.25	19.30	5.08	957	0.06	0.156
D1-92	8.07	9.79	7.27	1.21	6.11	10.87	7.19	19.62	5.75	799	0.14	0.157
D1-93	7.91	10.45	5.98	1.31	7.52	7.54	9.32	19.92	4.97	897	0.06	0.134
D1-94	7.82	10.62	6.24	1.22	6.37	8.69	9.22	19.78	5.66	818	0.06	0.077
D1-95	6.87	11.50	6.82	1.22	7.69	9.18	9.83	19.65	7.60	427	0.06	0.018

D1-96	8.83	**10.17**	**6.80**	**1.16**	**6.24**	**9.90**	7.92	19.64	5.98	426	0.152	0.034
D1-97	12.21	**11.44**	**10.03**	**3.86**	**8.92**	**13.68**	9.64	19.21	7.93	444	0.083	0.019
D1-98	12.85	**11.92**	**10.32**	**3.56**	**9.25**	**14.25**	10.22	19.24	8.64	439	0.362	0.006
D1-99	13.93	**11.86**	**11.31**	**6.18**	**9.83**	**14.62**	10.37	19.57	8.06	454	0.339	0.058
D1-100	12.10	**11.22**	**7.73**	**1.26**	**8.90**	**10.39**	8.50	19.36	8.52	470	0.12	0.062
D1-101	9.45	**10.84**	**7.01**	**1.37**	**6.76**	**9.95**	9.42	19.09	6.58	484	0.215	0.042
D1-102	10.50	**11.46**	**8.37**	**3.11**	**7.76**	**11.31**	10.01	19.29	7.51	494	0.075	0.014
D1-103	11.97	**11.12**	**7.56**	**1.74**	**6.92**	**10.80**	9.12	18.73	7.98	513	0.06	0.017
D1-104	7.40	**11.21**	**7.20**	**2.06**	**7.88**	**9.43**	9.98	19.27	6.86	495	0.06	0.013
D1-105	9.88	**10.86**	**8.99**	**3.88**	**7.57**	**12.25**	9.12	19.68	6.55	536	0.06	0.025
D1-106	8.39	**10.32**	**7.91**	**2.24**	**6.86**	**11.28**	8.38	19.33	6.11	555	0.06	0.04
D1-107	6.48	**10.67**	**6.08**	**1.89**	**7.09**	**7.67**	10.73	17.22	5.93	153	0.273	0.063
D1-108	9.38	**11.97**	**10.15**	**4.88**	**8.54**	**13.58**	12.14	18.83	6.88	140	0.15	0.026
D1-109	9.38	**12.62**	**9.80**	**4.26**	**8.92**	**13.00**	12.79	19.23	7.72	120.4	0.173	0.022
D1-110	11.60	**14.00**	**13.12**	**8.14**	**10.43**	**16.96**	14.29	19.49	9.76	190	0.125	0.026
D1-111	15.40	**14.66**	**12.00**	**5.85**	**10.07**	**16.05**	15.23	19.93	10.26	142	0.148	0.023
D1-112	13.29	**14.72**	**13.23**	**9.15**	**10.62**	**16.58**	15.14	19.99	10.46	152	0.311	0.036
D1-113	9.30	**12.25**	**9.63**	**2.17**	**9.45**	**13.44**	11.76	19.31	7.77	176	0.36	0.03
D1-114	12.60	**13.57**	**10.90**	**6.44**	**10.34**	**13.40**	13.83	19.63	8.95	142	0.259	0.054
D1-115	11.34	**13.85**	**11.85**	**5.55**	**9.90**	**15.97**	14.43	19.61	9.06	106.3	0.315	0.027
D1-116	8.81	**12.29**	**10.56**	**4.92**	**8.73**	**14.30**	12.14	19.07	7.62	115.5	0.226	0.029
D1-117	3.88	**11.33**	**6.81**	**2.36**	**7.63**	**8.63**	11.54	17.81	6.46	191	0.454	0.08
D2-1	18.28	**16.19**	**12.52**	**9.57**	**10.44**	**15.04**	18.90	19.96	10.40	74.8	0.126	0.031
D2-2	19.37	**16.43**	**12.65**	**9.97**	**10.47**	**15.08**	19.49	19.95	10.44	71.6	0.07	0.0115
D2-3	18.80	**16.29**	**12.70**	**9.88**	**10.41**	**15.26**	19.16	19.94	10.38	102	0.07	0.0115
D2-4	18.20	**16.22**	**12.55**	**9.65**	**10.48**	**15.02**	19.22	19.89	10.18	181	0.07	0.0115
D2-5	18.78	**16.26**	**12.53**	**9.58**	**10.50**	**15.02**	19.14	19.90	10.35	169	0.148	0.0115
D2-6	17.31	**15.80**	**12.61**	**9.51**	**10.34**	**15.31**	18.47	19.61	10.03	48.8	0.07	0.056
D2-7	18.40	**16.06**	**12.56**	**9.07**	**10.29**	**15.43**	18.56	19.91	10.45	194	0.07	0.0115
D2-8	18.42	**15.91**	**12.97**	**9.08**	**10.47**	**16.17**	18.61	19.54	10.23	652	0.07	0.0115
D2-9	5.94	**11.56**	**7.45**	**2.95**	**8.30**	**9.27**	11.26	19.44	6.29	920	0.169	0.0115
D2-10	16.81	**14.76**	**15.09**	**13.02**	**12.18**	**17.58**	15.39	19.85	10.41	274	0.07	0.0115
D2-11	19.44	**16.39**	**12.95**	**10.17**	**10.54**	**15.55**	19.35	19.92	10.47	277	0.186	0.0115
D2-12	13.42	**13.50**	**10.79**	**4.26**	**9.00**	**14.96**	13.04	19.45	9.79	839	0.07	0.03
D2-13	17.90	**14.78**	**13.34**	**10.35**	**10.86**	**16.08**	16.16	19.29	10.00	259	0.07	0.0115
D2-14	19.40	**15.98**	**12.85**	**10.17**	**10.53**	**15.35**	18.28	19.99	10.49	270	0.07	0.0115
D2-15	17.92	**15.21**	**13.08**	**8.16**	**10.47**	**16.86**	16.69	19.59	10.38	277	0.156	0.0115
D2-16	18.84	**15.87**	**12.62**	**9.68**	**10.50**	**15.15**	18.13	19.94	10.39	292	0.203	0.0115
D2-17	16.59	**13.87**	**11.86**	**8.27**	**10.10**	**14.54**	14.62	18.57	9.66	377	0.366	0.0115
D2-18	18.01	**16.15**	**12.59**	**9.85**	**10.30**	**15.10**	19.18	19.68	10.17	180	0.07	0.0115
D2-19	13.01	**15.04**	**8.92**	**4.40**	**10.30**	**10.48**	15.94	19.84	10.58	141	1.268	0.0115
D2-20	11.71	**12.61**	**10.06**	**4.34**	**8.63**	**13.63**	12.74	19.09	7.84	761	0.07	0.026
D2-21	16.46	**15.74**	**11.52**	**7.74**	**10.49**	**13.92**	17.48	20.00	10.71	471	0.07	0.0115
D2-22	19.05	**16.29**	**12.61**	**9.88**	**10.35**	**15.10**	19.06	20.00	10.46	177	0.07	0.071

D2-23	18.51	**16.29**	**12.61**	**9.89**	**10.43**	**15.06**	19.24	19.90	10.35	124.6	0.409	0.059
D2-24	19.05	**16.45**	**12.79**	**10.14**	**10.50**	**15.27**	19.52	19.93	10.45	47.5	0.07	0.0115
D2-25	18.80	**16.29**	**12.99**	**10.46**	**10.50**	**15.50**	19.20	19.83	10.43	29.6	0.07	0.0115
D2-26	17.88	**16.13**	**12.55**	**9.35**	**10.49**	**15.19**	18.36	19.79	10.96	47.7	0.217	0.0115
D2-27	19.42	**16.35**	**12.81**	**10.21**	**10.50**	**15.26**	19.35	19.90	10.39	370	0.07	0.0115
D2-28	17.39	**15.98**	**13.36**	**10.69**	**11.07**	**15.83**	18.39	19.90	10.44	51.2	0.07	0.0115
D2-29	4.13	**7.98**	**3.39**	**1.00**	**5.81**	**3.38**	10.40	10.50	3.40	1207	33.823	5.432
D2-30	8.45	**7.01**	**7.71**	**2.42**	**7.25**	**10.58**	6.21	15.35	1.97	1060	4.093	0.809
D2-31	4.30	**9.75**	**6.29**	**1.48**	**7.28**	**8.21**	7.87	17.69	6.22	1610	8.821	0.063
D2-32	13.09	**13.78**	**11.31**	**5.10**	**10.28**	**14.93**	14.34	19.86	8.81	459	0.561	0.054
D2-33	19.54	**15.77**	**13.42**	**10.48**	**10.54**	**16.33**	18.71	19.24	9.94	56.6	0.14	0.0115
D2-34	19.84	**17.35**	**12.87**	**10.40**	**10.50**	**15.30**	19.81	19.98	12.66	54.5	0.143	0.0115
D2-35	18.92	**16.26**	**12.91**	**10.14**	**10.52**	**15.49**	19.16	19.79	10.41	111	0.133	0.0115
D2-36	19.60	**16.53**	**13.00**	**10.67**	**10.59**	**15.37**	19.64	20.00	10.50	191	0.07	0.024
D2-37	10.43	**10.90**	**7.01**	**1.67**	**7.06**	**9.65**	7.27	19.03	9.23	709	0.277	0.053
D2-38	9.46	**10.25**	**6.73**	**1.00**	**7.21**	**9.35**	6.86	18.82	7.99	728	0.239	0.065
D2-39	10.58	**10.87**	**6.94**	**1.05**	**6.03**	**10.35**	5.87	19.79	10.22	694	0.07	0.027
D2-40	10.58	**11.02**	**6.98**	**1.05**	**6.10**	**10.39**	5.98	19.95	10.41	734	0.07	0.025
D2-41	10.66	**10.96**	**6.89**	**1.00**	**6.02**	**10.28**	5.79	19.96	10.44	747	0.07	0.029
D2-42	15.44	**12.99**	**11.14**	**5.42**	**10.04**	**14.55**	11.93	18.20	10.48	373	0.07	0.0115
D2-43	6.91	**8.41**	**8.53**	**1.00**	**8.98**	**12.08**	6.86	16.76	4.21	1510	9.243	1.091
D2-44	11.72	**11.65**	**8.72**	**1.74**	**9.30**	**11.92**	9.60	17.91	9.51	872	3.822	0.068
D2-45	11.93	**11.44**	**11.20**	**4.49**	**10.46**	**14.92**	10.20	17.75	8.35	629	2.013	0.035
D2-46	14.17	**13.46**	**10.85**	**6.55**	**9.39**	**13.72**	11.95	19.97	10.54	590	0.07	0.0115
D2-47	16.54	**14.25**	**14.05**	**11.20**	**11.54**	**16.72**	14.09	19.89	10.41	800	0.502	0.025
D2-48	11.44	**12.87**	**10.10**	**3.21**	**9.87**	**13.65**	11.88	19.02	9.61	1360	3.612	0.035
D2-49	13.16	**12.00**	**10.93**	**4.65**	**9.91**	**14.57**	9.20	19.39	9.92	1053	159.046	0.032
D3-1	14.43	**11.47**	**12.00**	**9.28**	**9.95**	**14.37**	14.86	15.04	5.03	370	1.35	0.06
D3-2	4.10	**6.71**	**3.67**	**1.27**	**6.63**	**3.39**	10.87	7.29	1.53	580	8.7	2.8
D3-3	3.71	**6.94**	**4.03**	**1.27**	**7.10**	**3.88**	10.55	8.01	2.05	570	3.25	0.44
D3-4	6.05	**10.28**	**4.42**	**1.29**	**5.94**	**5.24**	10.68	17.54	4.65	460	0.73	0.11
D3-5	8.32	**8.63**	**8.18**	**1.34**	**6.45**	**12.47**	10.91	14.10	2.11	420	9.9	1.6
D3-6	12.07	**12.66**	**15.31**	**14.50**	**12.22**	**17.26**	12.08	18.43	9.19	440	0.3	0.02
D3-7	7.23	**10.61**	**9.17**	**5.10**	**8.24**	**11.67**	10.77	18.01	5.16	630	0.6	0.15
D3-8	8.14	**9.92**	**9.54**	**4.44**	**8.43**	**12.64**	10.56	15.64	5.10	450	2.95	0.29
D3-9	12.02	**10.98**	**10.82**	**5.69**	**8.94**	**14.32**	12.13	16.53	5.71	180	0.3	0.03
D3-10	11.61	**10.50**	**10.78**	**5.19**	**9.73**	**14.11**	10.74	16.56	5.91	490	0.3	0.02
D3-11	8.29	**10.73**	**8.21**	**3.42**	**8.19**	**10.61**	10.73	17.84	5.64	490	0.3	0.02
D3-12	6.50	**9.58**	**5.56**	**1.47**	**6.44**	**7.17**	10.50	16.32	3.71	380	3.5	0.25
D3-13	17.19	**15.27**	**13.06**	**9.45**	**10.56**	**16.12**	17.09	19.55	10.14	340	0.3	0.02
D3-14	7.53	**10.11**	**8.13**	**3.99**	**7.78**	**10.37**	11.19	15.71	4.89	400	1.21	0.02
D3-15	4.37	**10.33**	5.98	1.60	7.05	7.64	11.55	15.85	5.00	280	0.73	0.04
D3-16	16.77	**15.94**	**10.74**	**7.00**	**8.92**	**13.52**	18.99	19.63	9.81	170	0.63	0.07
D3-17	17.95	**16.02**	**13.08**	**10.31**	**10.49**	**15.77**	18.71	19.83	10.21	140	0.78	0.03

D3-18	18.40	**15.96**	**13.26**	**10.77**	**10.84**	**15.71**	18.87	19.92	9.81	100	0.3	0.01
D3-19	18.02	**15.98**	**13.74**	**10.77**	**11.05**	**16.57**	18.87	19.96	9.83	87	0.3	0.02
D3-20	19.27	**16.18**	**13.14**	**10.55**	**10.59**	**15.71**	19.19	19.94	10.05	81	0.3	0.03
D3-21	7.93	**10.69**	**7.64**	**2.49**	**6.87**	**10.61**	10.85	18.41	4.99	460	1.12	0.22
D3-22	4.62	**8.96**	**8.69**	**3.41**	**7.65**	**11.86**	10.51	12.74	4.49	500	1.45	0.07
D3-23	13.41	**12.18**	**11.79**	**7.06**	**9.54**	**15.29**	12.60	18.91	6.89	500	2.07	0.15
D3-24	10.81	**10.29**	**10.18**	**4.60**	**9.01**	**13.56**	10.63	16.35	5.58	660	1.55	0.06
D3-25	12.85	**10.46**	**12.38**	**6.69**	**9.93**	**16.44**	11.67	16.14	5.02	520	1.15	1.9
D3-26	17.52	**14.43**	**13.10**	**10.36**	**10.47**	**15.78**	17.08	16.96	9.59	150	2.65	0.01
D3-27	14.90	**12.40**	**13.66**	**9.91**	**10.51**	**17.12**	14.20	17.66	6.58	190	1.35	0.04
D3-28	9.65	**11.29**	**10.56**	**6.29**	**8.74**	**13.60**	11.48	18.41	6.00	500	0.3	0.02
D3-29	15.52	**15.11**	**12.13**	**9.51**	**10.06**	**14.47**	17.09	19.67	9.60	660	0.3	0.11
D3-30	10.76	**10.27**	**13.02**	**7.43**	**9.19**	**17.72**	12.48	14.99	4.36	340	3.25	0.11
D3-31	12.65	**14.04**	**14.59**	**11.17**	**11.99**	**17.60**	17.69	18.37	6.78	170	0.97	0.23
D3-32	6.99	**10.56**	**8.06**	**2.80**	**7.29**	**11.08**	11.37	17.42	4.73	670	0.91	0.52
D3-33	10.47	**11.56**	**10.84**	**5.67**	**9.03**	**14.33**	12.88	16.42	6.58	210	1.85	0.02
D3-34	6.03	**10.56**	**4.99**	**1.32**	**6.24**	**6.20**	10.73	18.32	4.84	480	0.36	1.1
D3-35	5.61	**8.31**	**3.06**	**1.00**	**5.76**	**2.75**	10.44	13.81	1.96	440	2.55	1.3
D3-36	7.01	**9.85**	**7.56**	**2.87**	**7.40**	**9.98**	10.69	16.52	4.13	340	1.75	0.14
D3-37	3.61	**6.97**	**3.33**	**1.42**	**6.05**	**2.93**	10.71	8.29	1.74	520	6.1	1.3
D3-38	9.69	**8.64**	**9.95**	**3.48**	**7.84**	**14.23**	10.06	12.55	4.23	530	5.25	0.71
D3-39	6.86	**9.99**	**6.83**	**3.27**	**6.64**	**8.70**	10.71	16.18	4.75	350	4.24	0.33
D3-40	13.03	**10.00**	**11.59**	**5.82**	**10.35**	**15.09**	10.63	15.92	5.06	760	0.3	0.02
D3-41	7.04	**10.70**	**6.29**	**1.79**	**6.37**	**8.49**	10.49	18.65	5.26	640	1.73	0.14
D3-42	8.61	**10.43**	**9.04**	**4.50**	**7.89**	**11.88**	10.66	17.71	4.96	840	3.05	0.07
D3-43	16.41	**15.48**	**14.34**	**11.29**	**11.51**	**17.27**	18.10	19.64	9.52	120	0.44	0.02
D3-44	8.42	**10.44**	**5.06**	**1.90**	**6.47**	**5.93**	13.48	14.28	4.22	260	1.25	0.11
D3-45	8.09	**8.33**	**8.25**	**2.48**	**7.91**	**11.31**	10.54	12.44	2.86	620	6.85	2.4
D3-46	5.27	**7.07**	**4.02**	**1.64**	**6.14**	**4.16**	10.47	9.53	1.42	330	1.06	0.04
D3-47	2.61	**8.40**	**3.38**	**1.31**	**6.21**	**3.01**	10.56	11.71	3.57	970	10.2	4.4
D3-48	8.83	**9.18**	**9.64**	**4.57**	**9.03**	**12.49**	10.98	14.05	3.64	770	10.25	4.1
D3-49	11.48	**12.65**	**13.10**	**8.49**	**9.90**	**17.01**	15.01	16.98	6.86	370	0.3	0.02
D3-50	9.89	**11.70**	**9.32**	**3.39**	**8.05**	**12.92**	13.36	17.40	5.73	340	0.3	0.05
D3-51	10.36	**9.03**	**9.74**	**2.86**	**7.67**	**14.21**	10.75	14.51	3.15	520	2.25	0.77
D3-52	10.71	**9.69**	**10.40**	**4.22**	**8.26**	**14.55**	10.76	15.39	4.39	570	2.85	0.95
D3-53	12.88	**11.90**	**12.26**	**9.35**	**9.99**	**14.85**	14.90	15.70	5.75	190	1.15	0.04
D3-54	8.74	**8.55**	**8.56**	**1.57**	**6.25**	**13.21**	11.41	12.97	2.12	220	2.45	0.36
D3-55	4.16	**8.48**	**5.82**	**2.59**	**7.15**	**6.78**	10.36	13.13	3.00	600	5.65	0.1
D3-56	12.28	**10.69**	**13.20**	**7.66**	**9.82**	**17.66**	12.38	15.45	5.37	100	0.49	0.02
D3-57	9.33	**9.78**	**8.42**	**4.20**	**8.42**	**10.53**	10.61	15.05	5.06	360	3.45	0.02
D3-58	19.98	**16.13**	**12.90**	**10.48**	**10.50**	**15.30**	19.10	19.95	10.01	260	0.7	0.03
D3-59	12.82	**10.61**	**11.82**	**6.16**	**9.87**	**15.62**	11.40	16.11	5.78	260	1.14	0.16
D3-60	12.62	**11.67**	**11.91**	**9.40**	**9.95**	**14.14**	14.56	15.49	5.64	260	0.98	0.02
D4-1	8.15	**9.98**	**7.95**	**1.92**	**6.65**	**11.62**	11.08	16.80	3.86	707	1.95	0.14

D4-2	19.07	14.54	14.03	10.50	10.56	17.53	17.05	18.53	8.80	135.2	0.37	0.05
D4-3	6.61	9.16	7.27	2.46	7.60	9.51	11.49	13.20	3.62	326	1.73	0.28
D4-4	7.21	9.43	5.86	2.34	6.73	7.19	11.63	14.46	3.32	401	0.38	0.14
D4-5	18.79	14.25	11.37	5.69	9.20	15.30	16.60	19.48	7.83	211	0.3	0.07
D4-6	17.35	15.36	13.46	9.64	10.26	16.98	17.50	19.66	9.84	149.2	0.24	0.05
D4-7	12.06	10.82	11.21	4.40	9.30	15.57	11.80	16.74	5.49	563	3.68	1.42
D4-8	7.21	9.43	5.86	2.34	6.73	7.19	11.63	14.46	3.32	514	6.64	2.5
D4-9	13.20	11.29	11.71	8.58	9.81	14.22	14.53	14.93	4.99	392	2.58	0.11
D4-10	6.48	10.37	6.93	2.21	6.67	9.42	11.21	17.14	4.58	521	13.58	1.56
D4-11	5.10	6.99	5.32	1.84	8.23	5.60	10.41	8.92	1.69	512	8.06	0.99
D4-12	16.49	9.90	10.55	3.11	9.72	14.68	10.52	16.90	4.20	556	1.04	2.17
D4-13	14.42	10.69	10.58	5.15	9.48	13.85	11.52	16.48	5.59	157	1.96	0.04
D4-14	12.58	11.76	11.62	6.44	10.83	14.61	11.17	18.68	7.49	500	0.1	0.05
D4-15	12.34	9.44	9.31	2.45	7.04	13.87	10.92	15.83	3.19	465	3.38	0.2
D4-16	16.00	14.42	12.58	8.79	10.94	15.29	17.14	18.62	8.30	174.9	0.37	0.04
D4-17	13.68	11.13	11.95	6.80	9.95	15.52	12.92	16.10	5.52	199	1.26	0.05
D4-18	12.77	10.27	6.48	2.32	6.70	8.44	10.72	17.77	4.38	328	1.81	0.11
D4-19	14.60	11.10	12.01	6.05	9.58	16.20	12.20	15.54	6.68	247	0.48	0.04
D4-20	6.98	9.47	7.92	4.91	7.76	9.50	13.13	12.53	3.09	224	0.82	0.04
D4-21	5.61	8.84	7.19	2.98	8.32	8.73	10.98	13.05	3.39	361	1.47	0.12
D4-22	4.24	7.19	5.17	2.72	6.96	5.49	10.82	8.96	1.79	771	8.92	3.72
D4-23	8.29	9.40	8.42	4.73	7.92	10.52	10.55	15.68	3.60	449	4.17	0.16
D4-24	17.10	11.94	12.76	10.36	10.43	15.13	15.23	15.28	5.79	200	1.43	0.06
D4-25	13.30	11.22	11.87	8.65	9.73	14.55	14.75	14.51	4.83	217	1.77	0.18
D4-26	11.69	10.98	10.84	5.28	9.81	14.14	11.42	16.80	6.32	482	1.81	0.05
D4-27	15.33	9.59	10.20	4.14	8.89	13.88	10.63	15.97	3.85	438	1.73	0.17
D4-28	15.56	9.89	12.01	5.88	10.21	15.97	10.80	15.40	4.91	376	0.7	0.05
D4-29	13.56	9.94	11.24	6.23	9.87	14.43	10.59	15.73	5.05	1157	2.25	0.05
D4-30	8.46	9.71	7.68	3.27	7.66	9.88	10.58	15.57	4.53	327	2.1	0.13
D4-31	6.49	9.49	4.84	1.84	6.18	5.66	10.51	14.82	4.53	438	2.27	0.15
D4-32	11.35	11.30	10.63	4.43	9.49	14.30	12.92	16.74	5.57	201	0.27	0.03
D4-33	9.79	11.65	9.34	4.16	8.55	12.32	12.41	18.04	6.21	439	0.36	0.05
D4-34	11.13	12.73	8.26	3.28	7.85	10.95	13.80	18.67	7.27	294	0.09	0.06
D4-35	8.16	9.60	10.24	5.30	8.93	13.36	11.04	14.56	4.42	599	1.58	0.51
D4-36	16.35	14.98	12.56	8.28	10.50	15.72	17.81	18.99	8.87	149.6	0.06	0.04
D4-37	13.59	14.55	11.85	7.20	9.66	15.27	17.39	19.36	7.87	360	0.2	0.1
D4-38	12.95	10.09	11.55	5.60	10.31	15.15	10.67	15.37	5.65	512	2.09	1.59
D4-39	9.01	10.48	8.09	2.03	6.79	11.77	10.76	16.99	5.52	372	0.58	0.04
D5-1	2.37	6.67	4.92	1.07	6.54	6.05	6.95	12.60	2.10	1600	8.76	0.89
D5-2	4.18	5.69	5.96	1.12	7.64	7.55	5.28	11.13	2.26	1600	8.77	0.95
D5-3	8.45	8.10	8.31	5.45	7.11	10.34	8.29	12.17	4.99	760	2.72	0.097
D5-4	9.60	8.00	8.62	5.80	7.41	10.63	7.99	12.02	5.15	750	2.86	0.103
D5-5	11.43	10.08	10.41	6.88	9.12	12.83	10.56	15.64	5.57	580	2.31	0.12
D5-6	11.63	10.06	10.52	6.14	8.78	13.57	10.04	16.63	5.41	630	3.28	0.1

D5-7	9.04	**11.17**	**9.98**	**7.02**	**8.89**	**12.01**	12.27	15.26	7.00	370	0.72	0.005
D5-8	8.06	**10.24**	**7.93**	**3.92**	**7.72**	**10.04**	10.25	15.00	6.82	380	0.71	0.005
D5-9	16.69	**13.68**	**12.58**	**10.40**	**10.32**	**14.81**	14.02	18.25	10.02	410	2.74	0.005
D5-10	9.66	**7.18**	**7.90**	**4.92**	**6.16**	**10.25**	6.35	11.25	5.22	1030	2.79	0.15
D5-11	9.52	**7.33**	**7.97**	**4.91**	**6.23**	**10.36**	6.52	11.27	5.43	1210	2.54	0.18
D5-12	9.58	**9.21**	**9.88**	**4.71**	**9.38**	**12.72**	9.69	14.38	4.96	640	1.73	0.123
D5-13	11.31	**7.51**	**9.16**	**5.46**	**10.24**	**10.47**	6.65	11.38	5.73	680	6.12	0.042
D5-14	8.97	**8.19**	**8.71**	**4.29**	**9.28**	**10.65**	8.05	13.13	4.82	760	1.08	0.081
D5-15	10.44	**8.18**	**9.28**	**5.10**	**9.74**	**11.15**	7.81	12.27	5.68	760	0.99	0.085

Table A.5: Dataset 1 ecological classifications.

Ecological class	IPS	WAI-AVG (present study)	KBI-A (present study)	KBI-N (present study)	KBI-O (present study)	KBI-I (present study)	WAI-EC (present study)	WAI-DIN (present study)	WAI-OP (present study)
Eutrophic	34	14	57	116	66	12	35	0	91
Meso-eutrophic	44	55	51	1	51	46	40	0	26
Mesotrophic	34	47	9	0	0	49	38	2	0
Meso-oligotrophic	5	1	0	0	0	9	4	13	0
Oligotrophic	0	0	0	0	0	1	0	102	0

Table A.6: Dataset 2 ecological classifications.

Ecological class	IPS	WAI-AVG (present study)	KBI-A (present study)	KBI-N (present study)	KBI-O (present study)	KBI-I (present study)	WAI-EC (present study)	WAI-DIN (present study)	WAI-OP (present study)
Eutrophic	16	14	23	34	22	13	19	11	20
Meso-eutrophic	9	9	10	25	37	9	8	1	39
Mesotrophic	5	11	26	1	1	11	5	0	1
Meso-oligotrophic	5	25	1	0	0	26	4	2	0
Oligotrophic	25	1	0	0	0	1	24	46	0

Table A.7: Dataset 3 ecological classifications.

Ecological class	IPS	WAI-AVG (present study)	KBI-A (present study)	KBI-N (present study)	KBI-O (present study)	KBI-I (present study)	WAI-EC (present study)	WAI-DIN (present study)	WAI-OP (present study)
Eutrophic	27	12	24	45	34	14	0	3	49
Meso-eutrophic	11	33	19	14	25	10	36	2	11
Mesotrophic	12	6	16	1	1	18	12	11	0
Meso-oligotrophic	3	9	1	0	0	11	1	21	0
Oligotrophic	7	0	0	0	0	7	11	23	0

Table A.8: Dataset 4 ecological classifications.

Ecological class	IPS	WAI-AVG (present study)	KBI-A (present study)	KBI-N (present study)	KBI-O (present study)	KBI-I (present study)	WAI-EC (present study)	WAI-DIN (present study)	WAI-OP (present study)
Eutrophic	13	3	15	36	19	7	0	2	38
Meso-eutrophic	8	29	17	3	20	8	24	0	1
Mesotrophic	12	6	7	0	0	12	8	9	0
Meso-oligotrophic	4	1	0	0	0	11	2	17	0
Oligotrophic	2	0	0	0	0	1	5	11	0

Table A.9: Dataset 5 ecological classifications.

Ecological class	IPS	WAI-AVG (present study)	KBI-A (present study)	KBI-N (present study)	KBI-O (present study)	KBI-I (present study)	WAI-EC (present study)	WAI-DIN (present study)	WAI-OP (present study)
Eutrophic	5	9	8	14	9	2	9	0	14
Meso-eutrophic	9	5	6	1	6	8	4	4	1
Mesotrophic	0	1	1	0	0	5	2	7	0
Meso-oligotrophic	1	0	0	0	0	0	0	3	0
Oligotrophic	0	0	0	0	0	0	0	1	0

Appendix B: Ancillary tables for AMD index.

Table B.1: The optimum environmental conditions for 78 diatom taxa for EC, pH, sulphate, chloride and alkalinity, as well as the AMD and osmotic tolerance value and life-form.

Taxon	EC (µS/cm)	pH	Sulphate (mg/L)	Chloride (mg/L)	Alkalinity (mg/L)	AMD (tolerance or sensitivity)	Life-form	Osmotic pressure (tolerance or sensitivity)
Achnanthidium minutissimum (Kützing) Czarnecki	312.02	8.7	75.97	9.83	54.31	2	3	2
Achnanthidium standeri (Cholnoky) J.C. Taylor, E. Morales & Ector	178.7	8.33	35.87	6.91	37.55	2	3	2
Adlafia bryophila (Petersen) Lange-Bertalot	154.1	7.54	17.9	8.58	41.89	1	1	2

Amphora veneta Kützing	741.37	9.42	39.73	108.43	157.87	2	3	1
Aulacoseira granulata (Ehrenberg) Simonsen	231.45	8.36	33.12	12.56	56	2	3	1
Aulacoseira granulata var. *angustissima* (O.Müller) Simonsen	162.36	8.45	29.03	8.46	31.52	2	3	1
Brachysira neoexilis Lange-Bertalot	425.87	7.85	150.45	8.92	43.54	1	1	1
Caloneis bacillum (Grunow) Cleve	415.56	7.99	82.98	14.22	101.22	1	1	2
Cocconeis placentula Ehrenberg	190.36	8.64	10.79	14.07	58.55	2	3	2
Craticula buderi (Hustedt) Lange-Bertalot	1627.67	8.39	566.63	152.99	77.91	1	1	1
Craticula molestiformis (Hustedt) Mayama	860.46	9.39	47.62	123.66	180.99	1	1	1
Cyclotella meneghiniana Kützing	376.28	8.52	30.34	24.91	121.98	2	3	1
Diadesmis confervacea Kützing	277.47	8.17	29.57	20.44	75.08	2	3	1
Encyonema minutum (Hilse) D.G. Mann	176.05	8.23	19.02	8.58	52.04	2	1	2
Encyonopsis microcephala (Grunow) Krammer	1361.12	8.31	789.86	22.24	49.12	1.5	1	2
Encyonopsis minuta Krammer & E. Reichardt	630.98	8.6	262.35	12.89	73.3	1.5	1	2
Encyonopsis subminuta Krammer & Lange-Bertalot	219.96	8.52	33.45	9.92	61.23	1	1	2
Eolimna subminuscula (Manguin) Gerd Moser, Lange-Bertalot & Metzeltin	219.96	8.52	33.45	9.92	61.23	2	1	1
Epithemia adnata (Kützing) Brébisson	197.27	8.57	28.79	8.81	45.92	2	1	1
Eunotia bilunaris (Ehrenberg) Schaarschmidt	118.97	7.13	26.18	4.38	20.12	1	3	1
Eunotia minor (Kützing) Grunow	116.35	7.93	9.62	10.26	28.58	1	3	2
Fragilaria biceps (Kützing) Lange-Bertalot	373.64	8.67	61.58	23.42	97.91	2	3	2
Fragilaria capucina Desmazières	110.34	7.83	11.93	5.78	29.4	2	3	2
Fragilaria nanana Lange-Bertalo	225.74	8.53	14.27	13.32	70.82	2	3	2
Fragilaria tenera (WM Smith) Lange-Bertalot	277.79	8.06	63.4	8.57	76.77	2	3	2
Fragilaria ulna var. *acus*	263.88	8.39	27.97	16.76	78.61	2	3	2
Fragilaria vaucheriae (Kützing) J.B. Petersen	233.97	8.51	32.65	9.86	66.64	2	3	1
Frustulia crassinervia (Brébisson) Lange-Bertalot & Krammer	1099.6	3.45	893.73	11.3	4.1	1	2	2
Gomphonema aff. *gracile*	262.23	8.26	68.19	9.41	47.53	2	3	2
Gomphonema aff. *lagenula* Kützing	63.04	7.1	8.27	3.19	15.51	2	3	2
Gomphonema angustatum (Kützing)	719.16	8.74	54.91	147.5	72.44	2	3	2
Gomphonema exilissimum Lange-Bertalot	353.28	7.57	100.45	13.07	39.76	2	3	2
Gomphonema gracile Ehrenberg	171.21	8.12	28.44	7.62	44.32	2	3	2
Gomphonema lagenula Kützing	170.29	7.98	16.46	9.26	52.59	2	3	2
Gomphonema parvulum (Kützing) Kützing	226.17	7.76	33.45	13.69	54.53	2	3	1
Gomphonema parvulum f. *saprophilum* Lange-Bertalot & E.Reichardt	529.55	8.66	17.02	104.14	81.52	2	3	1
Gomphonema pseudoaugur Krammer	273.51	8.53	20	25	73.41	2	3	1

Gomphonema spp.	190.6	7.98	34.77	8.59	40.89	1	3	1
Mayamaea atomus (Kützing) Lange-Bertlaot	821.46	8.61	53.83	128.19	157.94	1	1	1
Navicula arvensis var. *maior* Lange-Bertalot	539.45	9.45	25.57	72.12	131.35	2	1	1
Navicula cryptocephala Kützing	252.38	8.64	34.6	13.14	74.12	2	1	1
Navicula radiosa Kützing	119.67	8.2	6.59	4.41	44.76	2	1	1
Navicula recens (Lange-Bertalot) Lange-Bertalot	454.16	8.39	79.25	30.65	108.15	2	1	1
Navicula rostellata Kützing	258.42	8.43	49.08	8.98	62.01	2	1	1
Navicula schroeteri Meister	436.97	7.68	103.82	16.68	88.29	2	1	1
Navicula spp.	286.78	8.43	31.3	16	88.26	2	1	2
Navicula trivialis Lange-Bertalot	264.17	8.42	20.8	9.41	94.95	2	1	1
Navicula veneta Kützing	1069.75	8.5	353.02	89.54	82.81	1	1	1
Navicula zanonii Hustedt	255.84	8.61	33.59	13.02	76.15	2	1	2
Nitzschia acidoclinata Lange-Bertalot	244.25	7.69	82.73	5.9	26.4	1	1	2
Nitzschia amphibia Grunow	832.16	8.94	113.84	91.89	176.41	2	1	1
Nitzschia archibaldii Lange-Bertalot	250.39	8.14	43.36	11.33	63.49	2	1	1
Nitzschia capitellata Hustedt	2334.36	3.31	1438.43	12.42	2.81	1	1	1
Nitzschia clausii Hantzsch	403.33	7.18	142.48	6.29	45.26	1	1	1
Nitzschia draveillensis Coste & Ricard	326.1	7.96	31.73	24.92	94.27	2	1	1
Nitzschia filiformis W Smith	278.6	8.69	29.81	12.12	90.59	2	2	1
Nitzschia fonticola Grunow	437.48	7.97	106.01	16.4	83.29	1	1	1
Nitzschia frustulum (Kützing) Grunow	183.61	8.2	19.75	6.25	61.82	2	1	1
Nitzschia liebertruthii Rabenhorst	283.4	8.14	44.09	14.15	77.96	1	1	1
Nitzschia linearis var. *subtilis* Hustedt	222.31	8.35	30.14	10.82	60.89	2	1	1
Nitzschia linearis W. Smith	354.71	8.66	50.38	8.8	104.54	2	1	1
Nitzschia microcephala Grunow	670.34	8.63	80.44	69.4	168.6	2	1	1
Nitzschia nana Grunow	274.53	8.43	50.28	15.33	63.54	1	1	2
Nitzschia palea (Kützing) W. Smith	483.7	8.47	72.12	40.29	101.6	1	1	1
Nitzschia pura Hustedt	1749.83	8.1	787.4	45.69	66.61	1	1	2
Nitzschia radicula Hustedt	165.87	8.46	17.21	7.59	51.99	2	1	1
Nitzschia recta Hantzsch	86.23	7.61	10.9	2.75	25.94	2	1	1
Nitzschia reversa W. Smith	265.55	8.26	48.84	10.9	50.76	1	1	1
Nitzschia spp.	330.79	8.4	86.33	10.68	62.99	1	1	1
Pinnularia acrosphaeria W Smith	175.55	8.09	18.64	10.36	48.62	1	1	2
Pinnularia divergens W Smith	258.5	7.94	48.56	6.04	79.6	1	1	2
Pinnularia subbrevistriata Krammer	399.76	8.81	27.43	46.65	97.82	2	1	1
Pinnularia subcapitata Gregory	72.78	6.99	10.03	4.04	15.94	1	1	2
Pinnularia viridiformis Krammer	228.01	7.44	44.74	9.57	65.18	1	1	2
Planothidium rostratum (Østrup) Lange-Bertalot	261.43	8.2	27.44	14.79	77.86	2	3	1
Rhopalodia gibba (Ehrenberg) O.Müller	341.48	8.61	72.57	10.79	83.97	1	1	2

Sellaphora pupula (Kützing) Mereschkovsky	306.67	8.46	28.51	24.6	81.76	2	1	1
Sellaphora seminulum (Grunow) D.G. Mann	234.33	7.45	50.22	8.94	46.93	1	1	1

Table B.2: Calculated index score per site seasons. The MMI reversed score is calculated according to the methodology, with scores for EC and sulphate reversed, while the MMI score does not include reversing EC and sulphate scores.

Sample ID	Site	Disturbance level observed	MMI reversed	MMI Weighted
Mp4	1	2	78.22	89.11
Mp5	2	1	78.97	88.76
Mp11.2	3	1	65.29	74.49
Mp11.4	5	1	83.56	91.18
Mp13	6	1	80.44	88.71
Mp15	7	1	85.48	99.75
Mp16	8	2	91.09	94.39
MPW2	9	1	69.01	90.67
MPW3	10	1	74.80	93.43
MPW4	11	1	73.58	90.34
MPW4.2	12	1	79.43	89.70
MPW5	13	1	69.94	90.79
Vosstoffel	14	2	70.06	87.66
MP Ivan	15	1	76.35	89.58
Goedgevonden2/R	16	3	27.71	40.77
Goedgevonden1/L	17	3	11.87	18.10
Kendal PS	18	1	67.86	89.83
Bezhoek1	19	1	78.40	99.98
Bezhoek2	20	1	59.09	85.86
Bezhoek3	21	1	81.00	92.54
Jansen Wetland	22	1	83.34	91.12
KZN2	23	1	84.76	92.46
KZN3	24	3	73.96	90.11
KZN5	25	3	32.35	38.18
KZN6	26	1	77.66	84.22
KZN8	27	1	62.70	85.77
KZN9	28	1	100.00	96.54
KZN12	29	1	65.45	89.40
KZN17	30	3	81.63	87.40

KZN18	31	1	68.57	89.16
KZN19	32	1	72.95	95.72
KZN21	33	1	74.05	89.08
KZN36	34	3	32.80	49.29
KZN37	35	1	72.05	91.20
KZN40	36	1	80.39	90.60
KZN42	37	2	71.13	87.74
KZN47	38	1	75.60	87.84
KZN48	39	3	29.72	44.88
KZN51	40	3	62.57	86.47
KZN54	41	1	77.10	87.51
KZN56	42	1	75.78	91.00
KZN57	43	1	76.93	89.01
KZN58	44	1	79.07	100.00
WS1	45	2	64.90	87.40
WS2	46	1	64.71	88.17
Mp4	47	3	73.59	90.41
Mp5	48	1	71.58	91.04
Mp11.2	49	1	71.19	89.70
Mp11.3	50	1	71.87	91.86
Mp11.4	51	1	71.42	90.36
Mp13	52	1	71.10	92.48
Mp15	53	1	78.34	89.86
Mp16	54	1	87.77	93.49
MPW2	55	1	71.21	90.10
MPW3	56	1	69.89	91.01
MPW4	57	1	66.36	90.06
MPW4.2	58	1	71.82	91.87
MPW5	59	1	70.10	92.22
Vosstoffel	60	1	71.07	87.06
MP Ivan	61	1	79.64	87.92
Goedgevonden2/R	62	3	29.40	43.62
Goedgevonden1/L	63	3	0.00	0.00
Kendal PS	64	1	63.24	86.12
Jansen Wetland	65	1	74.69	92.82

KZN2	66	1	93.16	91.24
KZN3	67	1	68.60	88.65
KZN5	68	3	41.79	52.89
KZN6	69	1	52.30	66.27
KZN8	70	1	62.76	82.65
KZN9	71	1	81.46	99.45
KZN12	72	1	63.88	88.69
KZN17	73	3	68.31	86.58
KZN18	74	1	69.71	89.83
KZN19	75	1	71.21	92.22
KZN21	76	1	74.53	86.04
KZN36	77	4	40.24	59.93
KZN37	78	1	70.32	91.49
KZN40	79	1	79.17	90.61
KZN42	80	2	59.94	81.32
KZN47	81	1	78.89	90.09
KZN48	82	3	9.11	13.15
KZN54	83	1	78.55	88.70
KZN56	84	2	72.22	93.12
KZN57	85	1	78.79	93.66
KZN58	86	1	71.56	94.23
WS1	87	1	67.50	88.32
WS2	88	1	68.54	90.62

Table B.3: Water quality measurements for the five environmental variables used to create the index scores, with observed disturbance levels.

Sample ID	Site nr	Sulphate (mg/L)	Chloride (mg/L)	Alkalinity (mg/L)	EC (mS/cm)	pH	Disturbance level
Mp4	1	203	12	8.4	54	6.8	2
Mp5	2	30	19	43	23	9.6	1
Mp11.2	3	6.6	13	39	14	9.4	1
Mp11.4	5	8.1	24	52	21	9.6	1
Mp13	6	30	7	35	17	7.5	1
Mp15	7	12	3	7.9	6	6.7	1
Mp16	8	81	30	167	66	10	2
MPW2	9	17	10	31	15	7.9	1

MPW3	10	28	38	138	43	8.1	1
MPW4	11	38	5.6	34	19	8	1
MPW4.2	12	45	8.1	69	27	9	1
MPW5	13	29	9.4	80	23	8.9	1
Vosstoffel	14	29	88	143	62	9.7	2
MP Ivan	15	22	3.3	29	12	8.8	1
Goedgevonden2/R	16	497	2.3	1	108	3	3
Goedgevonden1/L	17	860	2	1	165	3.4	3
Kendal PS	18	3.8	2	48	10	7.9	1
Bezhoek1	19	2.3	2	18	5	7.4	1
Bezhoek2	20	10	2.6	10	6	6.9	1
Bezhoek3	21	3.1	3.3	6.6	3	6.8	1
Jansen Wetland	22	6.2	16	29	14	8.6	1
KZN2	23	5.3	245	70	92	9.4	1
KZN3	24	113	215	223	130	8.5	3
KZN5	25	1468	35	35	265	8.6	3
KZN6	26	6	2	62	14	9.6	1
KZN8	27	21	2.5	55	17	9.4	1
KZN9	28	2.5	4	107	23	9.5	1
KZN12	29	0.6	2	8.2	3	6.9	1
KZN17	30	309	8.1	69	74	9.3	3
KZN18	31	8.1	2	39	10	8.4	1
KZN19	32	10	2	4	4	6.3	1
KZN21	33	28	12	96	29	9.8	1
KZN36	34	3474	4.9	1	475	2.8	3
KZN37	35	10	5.9	46	14	7.8	1
KZN40	36	8.2	6.4	34	12	8.7	1
KZN42	37	315	12	27	72	8.7	2
KZN47	38	16	14	135	33	8.8	1
KZN48	39	1159	33	1	195	4.2	3
KZN51	40	1374	27	76	265	8.2	3
KZN54	41	16	7.3	88	22	9.2	1
KZN56	42	17	4.5	45	15	8.9	1
KZN57	43	19	2.7	32	12	9	1
KZN58	44	9.6	3.2	20	8	7.1	1
WS1	45	141	9.9	1.4	37	4.7	2
WS2	46	8	4.9	34	11	8.5	1
Mp4	47	245	15	256	86	8.2	3
Mp5	48	37	26	91	35	9.1	1
Mp11.2	49	7.1	24	85	27	8.7	1
Mp11.3	50	8.2	3	35	10	8.1	1
Mp11.4	51	6.7	18	94	25	8.7	1
Mp13	52	18	13	66	22	8.7	1
Mp15	53	43	6.6	17	17	7.8	1

Mp16	54	58	77	325	98	9.5	1
MPW2	55	17	34	143	41	8.4	1
MPW3	56	9.1	25	200	45	8.7	1
MPW4	57	51	9.1	52	25	7.9	1
MPW4.2	58	38	8	84	27	8.8	1
MPW5	59	29	16	98	31	8.6	1
Vosstoffel	60	17	138	236	90	9.3	1
MP Ivan	61	39	6.7	66	25	9.6	1
Goedgevonden2/R	62	466	6.3	1	90	3	3
Goedgevonden1/L	63	650	3.1	1	142	3.1	3
Kendal PS	64	3.3	2	155	31	7.7	1
Jansen Wetland	65	5.8	18	27	14	8.7	1
KZN2	66	7.4	325	97	124	9	1
KZN3	67	45	111	186	82	8.7	1
KZN5	68	1160	62	52	245	7.9	3
KZN6	69	5.2	2	102	21	9.1	1
KZN8	70	16	3.3	100	24	8	1
KZN9	71	8.9	12	82	22	8.8	1
KZN12	72	3.9	3.2	15	6	7	1
KZN17	73	575	13	55	210	8	3
KZN18	74	7.6	2	49	12	7.7	1
KZN19	75	23	6.3	45	15	8.6	1
KZN21	76	14	6.3	87	20	7.6	1
KZN36	77	3534	8.7	1	740	2.5	4
KZN37	78	6.5	3.4	50	12	8.8	1
KZN40	79	4.2	8.5	46	13	9.1	1
KZN42	80	141	13	67	58	8.7	2
KZN47	81	2.4	21	86	30	9.2	1
KZN48	82	1406	33	1	21	3.6	3
KZN54	83	12	11	84	21	9.3	1
KZN56	84	97	8	28	30	7.2	2
KZN57	85	23	4.7	38	14	9.4	1
KZN58	86	11	6.8	28	11	8.9	1
WS1	87	8.7	7.6	40	13	8.5	1
WS2	88	3.6	2.7	37	9	8	1

Table B.4: Wetland types for sites sampled with corresponding disturbance levels and GPS coordinates.

Site nr	Sample ID	Season	Type	Abbreviation	Disturbance level	Ref/ Dist	Location
1	Mp4	DRY	Seep	S	2	D	-25.9512356424, 29.156113379

2	Mp5	DRY	Seep	S	1	R	-25.9750832714, 29.150626362
3	Mp11.2	DRY	Channelled valley-bottom wetland	CVBW	1	R	-25.8151519447, 29.7950744621
5	Mp11.4	DRY	Seep	S	1	R	-25.8596233396, 29.810679921
6	Mp13	DRY	Seep	S	1	R	-26.1177090148, 29.0118737622
7	Mp15	DRY	Wetland Flat	WF	1	R	-26.0785114235, 29.0206846542
8	Mp16	DRY	Wetland Flat	WF	2	D	-26.1021753466, 28.9964742911
9	MPW2	DRY	Channelled valley-bottom wetland	CVBW	1	R	-26.291683, 29.084064
10	MPW3	DRY	Channelled valley-bottom wetland	CVBW	1	R	-26.268247, 29.00955
11	MPW4	DRY	Seep	S	1	R	-26.206667, 29.012133
12	MPW4.2	DRY	Seep	S	1	R	-26.206667, 29.012133
13	MPW5	DRY	Channelled valley-bottom wetland	CVBW	1	R	-26.157328, 28.94608
14	Vosstoffel	DRY	Wetland Flat	WF	2	D	-26.41675, 29.13209
15	MP Ivan	DRY	Channelled valley-bottom wetland	CVBW	1	R	-26.0958556, 29.0088861
16	Goedgevonden2/R	DRY	Channelled valley-bottom wetland	CVBW	3	D	-26.060364, 29.0710583
17	Goedgevonden1/L	DRY	Channelled valley-bottom wetland	CVBW	3	D	-26.0590028, 29.0662611
18	Kendal PS	DRY	Depression	D	1	R	-26.0622196501, 28.9540035389

19	Bezhoek1	DRY	Channelled valley-bottom wetland	CVBW	1	R	-25.7425111, 29.3310778
20	Bezhoek2	DRY	Channelled valley-bottom wetland	CVBW	1	R	-25.7461611, 29.3300417
21	Bezhoek3	DRY	Unchanneled valley-bottom wetland	UVBW	1	R	-25.7446056, 29.324675
22	Jansen Wetland	DRY	Unchanneled valley-bottom wetland	UVBW	1	R	-25.6572333, 29.5670361
23	KZN2	DRY	Seep	S	1	R	-27.8542994401, 30.1679377028
24	KZN3	DRY	Depression	D	3	D	-27.639516997, 30.1268284946
25	KZN5	DRY	Floodplain wetland	FPW	3	D	-27.6098621623, 30.1494935529
26	KZN6	DRY	Channelled valley-bottom wetland	CVBW	1	R	-27.6142183052, 30.1140342803
27	KZN8	DRY	Channelled valley-bottom wetland	CVBW	1	R	-27.6117969292, 30.1331093533
28	KZN9	DRY	Valley head seep	VHS	1	R	-27.6294036438, 30.1350810576
29	KZN12	DRY	Wetland Flat	WF	1	R	-27.8819842443, 30.1553308577
30	KZN17	DRY	Wetland Flat	WF	3	D	-27.8515660356, 30.1091353401
31	KZN18	DRY	Seep	S	1	R	-27.8523252468, 30.0987325487
32	KZN19	DRY	Seep	S	1	R	-27.8529105653, 30.1002134047
33	KZN21	DRY	Seep	S	1	R	-27.8563880915, 30.0977134068
34	KZN36	DRY	Valley head seep	VHS	3	D	-28.0564937359, 30.1319930023
35	KZN37	DRY	Wetland Flat	WF	1	R	-28.0660219083, 30.1261468652

36	KZN40	DRY	Channelled valley-bottom wetland	CVBW	1	R	-28.0838647893, 30.1494043745
37	KZN42	DRY	Channelled valley-bottom wetland	CVBW	2	D	-28.0475199785, 30.1680241634
38	KZN47	DRY	Seep	S	1	R	-28.0927716877, 30.1869078163
39	KZN48	DRY	Unchanneled valley-bottom wetland	UVBW	3	D	-28.0608365398, 30.0144388433
40	KZN51	DRY	Channelled valley-bottom wetland	CVBW	3	D	-28.0654720064, 29.9698637072
41	KZN54	DRY	Channelled valley-bottom wetland	CVBW	1	R	-28.0821613332, 30.013674451
42	KZN56	DRY	Seep	S	1	R	-28.0430131659, 30.0190240659
43	KZN57	DRY	Seep	S	1	R	-28.0390809322, 30.017521668
44	KZN58	DRY	Seep	S	1	R	-28.0423214149, 30.0005792695
45	WS1	DRY	Floodplain wetland	FPW	2	D	-27.3643316027, 30.1275673504
46	WS2	DRY	Channelled valley-bottom wetland	CVBW	1	R	-27.3318655572, 30.1586151225
47	Mp4	WET	Seep	S	3	D	-25.9512356424, 29.156113379
48	Mp5	WET	Seep	S	1	R	-25.9750832714, 29.150626362
49	Mp11.2	WET	Channelled valley-bottom wetland	CVBW	1	R	-25.8151519447, 29.7950744621
50	Mp11.3	WET	Channelled valley-bottom wetland	CVBW	1	R	-25.8463981331, 29.801032113
51	Mp11.4	WET	Seep	S	1	R	-25.8596233396, 29.810679921

52	Mp13	WET	Seep	S	1	R	-26.1177090148, 29.0118737622
53	Mp15	WET	Wetland Flat	WF	1	R	-26.0785114235, 29.0206846542
54	Mp16	WET	Wetland Flat	WF	1	R	-26.1021753466, 28.9964742911
55	MPW2	WET	Channelled valley-bottom wetland	CVBW	1	R	-26.291683, 29.084064
56	MPW3	WET	Channelled valley-bottom wetland	CVBW	1	R	-26.268247, 29.00955
57	MPW4	WET	Seep	S	1	R	-26.206667, 29.012133
58	MPW4.2	WET	Seep	S	1	R	-26.206667, 29.012133
59	MPW5	WET	Channelled valley-bottom wetland	CVBW	1	R	-26.157328, 28.94608
60	Vosstoffel	WET	Wetland Flat	WF	1	R	-26.41675, 29.13209
61	MP Ivan	WET	Channelled valley-bottom wetland	CVBW	1	R	-26.0958556, 29.0088861
62	Goedgevonden2/R	WET	Channelled valley-bottom wetland	CVBW	3	D	-26.060364, 29.0710583
63	Goedgevonden1/L	WET	Channelled valley-bottom wetland	CVBW	3	D	-26.0590028, 29.0662611
64	Kendal PS	WET	Depression	D	1	R	-26.09913, 28.96575
65	Jansen Wetland	WET	Unchanneled valley-bottom wetland	UVBW	1	R	-25.6572333, 29.5670361
66	KZN2	WET	Seep	S	1	R	-27.8542994401, 30.1679377028
67	KZN3	WET	Depression	D	1	R	-27.639516997, 30.1268284946
68	KZN5	WET	Floodplain wetland	FPW	3	D	-27.6098621623, 30.1494935529

69	KZN6	WET	Channelled valley-bottom wetland	CVBW	1	R	-27.6142183052, 30.1140342803
70	KZN8	WET	Channelled valley-bottom wetland	CVBW	1	R	-27.6117969292, 30.1331093533
71	KZN9	WET	Valley head seep	VHS	1	R	-27.6294036438, 30.1350810576
72	KZN12	WET	Wetland Flat	WF	1	R	-27.8819842443, 30.1553308577
73	KZN17	WET	Wetland Flat	WF	3	D	-27.8515660356, 30.1091353401
74	KZN18	WET	Seep	S	1	R	-27.8523252468, 30.0987325487
75	KZN19	WET	Seep	S	1	R	-27.8529105653, 30.1002134047
76	KZN21	WET	Seep	S	1	R	-27.8563880915, 30.0977134068
77	KZN36	WET	Valley head seep	VHS	4	D	-28.0564937359, 30.1319930023
78	KZN37	WET	Wetland Flat	WF	1	R	-28.0660219083, 30.1261468652
79	KZN40	WET	Channelled valley-bottom wetland	CVBW	1	R	-28.0838647893, 30.1494043745
80	KZN42	WET	Channelled valley-bottom wetland	CVBW	2	D	-28.0475199785, 30.1680241634
81	KZN47	WET	Seep	S	1	R	-28.0927716877, 30.1869078163
82	KZN48	WET	Unchanneled valley-bottom wetland	UVBW	3	D	-28.0608365398, 30.0144388433
83	KZN54	WET	Channelled valley-bottom wetland	CVBW	1	R	-28.0821613332, 30.013674451
84	KZN56	WET	Seep	S	2	D	-28.0430131659, 30.0190240659
85	KZN57	WET	Seep	S	1	R	-28.0390809322, 30.017521668

86	KZN58	WET	Seep	S	1	R	-28.0423214149, 30.0005792695
87	WS1	WET	Floodplain wetland	FPW	1	R	-27.3643316027, 30.1275673504
88	WS2	WET	Channelled valley-bottom wetland	CVBW	1	R	-27.3318655572, 30.1586151225